CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/11, C07K 14/00, C07H 21/04, C12Q 1/68, A61K 48/00, A01K 67/00

A1

(11) International Publication Number: WO 97/08308

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/US96/13896

(22) International Filing Date: 30 August 1996 (30.08.96)

(30) Priority Data:
60/003,030 31 August 1995 (31.08.95) US

(71) Applicants: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02110 (US). University College London [GB/GB]; Brook House, 2-16 Torrington Place, London WC1E 7HN (GB). LEIDEN UNIVERSITY [NL/NL]; Stationsweg 46, NL-Leiden (NL).

(72) Inventors: LERNER, Terry, J.; Massachusetts General Hospital, Charlestown, MA 02129 (US). TASCHNER, Peter, E., M.; Medical Genetic Center South-West Netherlands-MGC, P.O. Box 9503, NL-2300 RA Leiden (NL). BREUNING, Martijn, H.; Medical Genetics Center South-West Netherlands-MGC, P.O. Box 9503, NL-2300 RA Leiden (NL). GUSELLA, James, F.; Massachusetts General Hospital, Charlestown, MA 02129 (US). MOLE, Sara, E.; University College London Medical School, The Rayne Institute, London (GB). GARDINER, Mark, R.; University College London Medical School, The Rayne Institute, London (GB).

(74) Agents: MYERS, Paul, Louis et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, IS, JP, NO, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: BATTEN DISEASE GENE

(57) Abstract

A substantially pure nucleic acid which encodes a Batten disease polypeptide.

* (Referred to in PCT Gazette No. 21/1997, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

BATTEN DISEASE GENE

Government Funding

This invention was made with government support from the National Institutes of Health grants NS32009, NS24279, NS30152 and NS28722. The government has certain rights in the invention.

Field of the Invention

The invention relates to the Batten disease gene, Batten disease polypeptides, and methods using these and other related compounds.

Background of the Invention

The neuronal ceroid lipofuscinoses (NCLs) are a group of inherited neurodegenerative disorders characterized by the accumulation of autofluorescent lipopigments (ceroid and lipofuscin) in neurons and other cell types (Dyken et al. (1988) *Am. J. Med. Genet Suppl.* 6:69-84). At least five subtypes are recognized, based on age of onset, clinico-pathological features and chromosomal location. Inheritance is autosomal recessive for the childhood onset forms which include: infantile (*CLN1*; Haltia-Santavuori disease, MIM256730), classical late-infantile (*CLN2*; Jansky-Bielschowsky disease, MIM204500), juvenile (*CLN3*; Batten or Spielmeyer-Vogt-Sjogren disease, MIM304200), and Finnish variant late-infantile (*CLN5*; MIM256731). The primary biochemical defects in these disorders are not known.

Batten disease, the juvenile onset form of NCL, is the most common neurodegenerative disorder of childhood. Its incidence is estimated at up to 1/25,000 births (Zeman W. (1974) *J. Neuropathol. Exp. Neurol.* 33:1-12), with an increased prevalence in the North European population. Clinical onset begins with visual failure between the age of 5 and 10 years. Seizures and mental deterioration ensue with relentless decline to death usually in the second or third decade. Diagnostic criteria include the presence in many cell types of inclusions which appear as so-called "fingerprint profiles" on electron-microscopy (Wisniewski et al. (1988) *Am. J. Med. Genet. Suppl.* 5:17-46). The major protein component of these abnormal deposits is subunit 9 of mitochondrial ATPase (Palmer et al (1992) *Am. J. Med. Genet* 42:561-567, although the genetic defect does not lie in a gene encoding this 75 amino acid protein (Dyer et al. (1993) *Biochem. J.*293:51-64; Yan et al (1994) *Genomics*24:375-377.

Summary of the Invention

The inventors have identified and cloned the gene responsible for Batten disease, hereafter referred to as "the Batten disease gene." The gene is located on human chromosome 16p12.1 and encodes a polypeptide having a predicted 438 amino acid sequence. hereafter referred to as "a Batten disease polypeptide".

Accordingly, the invention features a polypeptide, e.g., a recombinant polypeptide or substantially pure preparation of a polypeptide, the sequence of which includes, or is, the sequence of a Batten disease polypeptide, e.g., the sequence shown in SEQ ID NO: 2 or SEQ ID NO: 19. The invention also features fragments and analogs preferably having at least one biological activity (as defined herein) of a Batten disease polypeptide.

In a preferred embodiments the polypeptide is a mammalian, e.g., a human or a rodent, e.g., a mouse or a rat, polypeptide.

In preferred embodiments: the polypeptide has at least one biological activity, e.g., it reacts with an antibody, or antibody fragment, specific for a Batten disease polypeptide; the polypeptide includes an amino acid sequence at least 60%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence from SEQ ID NO: 2 or SEQ ID NO: 19; the polypeptide includes an amino acid sequence more than 85% homologous to an amino acid sequence from SEQ ID NO: 2 or SEQ ID NO: 19; the polypeptide includes an amino acid sequence essentially the same as an amino acid sequence in SEQ ID NO: 2 or SEQ ID NO: 19; the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the polypeptide includes at least 5, preferably at least 10, more preferably at least 20, most preferably at least 50, 100, or 150 contiguous amino acids from SEQ ID NO: 2 or SEQ ID NO: 19; the polypeptide is preferably at least 10, but no more than 100, amino acids in length, and contains one, two, three or more phosphorylation sites; the Batten disease polypeptide is either, an agonist or an antagonist, of a biological activity of a naturally occurring Batten disease polypeptide.

In preferred embodiments: the Batten disease polypeptide is encoded by the nucleic acid sequence of SEQ ID NO: 1 or SEQ ID NO: 18, or by a nucleic acid having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with the nucleic acid of SEQ ID NO: 1; the polypeptide is encoded by a nucleic acid having more than 82% homology with the nucleic acid of SEQ ID NO: 1 or SEQ ID NO: 18. For example, the Batten disease polypeptide can be encoded by a nucleic acid sequence which differs from a nucleic acid sequence of SEQ ID NO: 1 or SEQ ID NO: 18 due to degeneracy in the genetic code.

In a preferred embodiments the nucleic acid encoding the Batten disease polypeptide is a mammalian, e.g., a human or a rodent, e.g., a mouse or a rat, nucleic acid.

In a preferred embodiment the Batten disease polypeptide is an agonist of a naturally-occurring mutant or wild type Batten disease polypeptide (e.g., a polypeptide

having an amino acid sequence shown in SEQ ID NO: 2 or SEQ ID NO: 19). In another preferred embodiment, the polypeptide is an antagonist which, for example, inhibits an undesired activity of a naturally-occurring Batten disease polypeptide (e.g., a mutant polypeptide).

In preferred embodiments, the Batten disease polypeptide includes amino acid residues 155-226 of SEQ ID NO: 2 and/or residues 255-352 of SEQ ID NO: 2.

In a preferred embodiment, the Batten disease polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence in SEQ ID NO: 2 or SEQ ID NO: 19. The differences, however, are such that the Batten disease polypeptide exhibits at least one biological activity of a Batten disease polypeptide, e.g., the Batten disease polypeptide retains a biological activity of a naturally occurring Batten disease polypeptide.

In preferred embodiments the Batten disease polypeptide includes a Batten disease polypeptide sequence, as described herein, as well as other N-terminal and/or C-terminal amino acid sequences.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence from SEQ ID NO: 2 or SEQ ID NO: 19, fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence from SEQ ID NO: 2 or SEQ ID NO: 19.

In yet other preferred embodiments, the Batten disease polypeptide is a recombinant fusion protein having a first Batten disease polypeptide portion and a second polypeptide portion having an amino acid sequence unrelated to a Batten disease polypeptide. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

In a preferred embodiment, the Batten disease polypeptide is a fragment or analog of a naturally occurring Batten disease polypeptide which inhibits reactivity with antibodies, or $F(ab')_2$ fragments, specific for a naturally occurring Batten disease polypeptide.

In a preferred embodiment, the Batten disease polypeptide includes a leader sequence, e.g., an N-terminal sequence responsible for secretion of the polypeptide from a cell in which it is expressed, or other sequence which is not present in the mature protein. In another preferred embodiment, the Batten disease polypeptide, e.g., the polypeptide of SEQ ID NO: 2 or SEQ ID NO: 19, lacks a leader sequence, e.g., an N-terminal sequence responsible for secretion of the polypeptide from a cell in which it is expressed, or other sequence which is not present in the mature protein.

In a preferred embodiment, the Batten Disease polypeptide has a molecular weight of about 48 kDa.

Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

The invention includes an immunogen which includes an active or inactive Batten disease polypeptide, or an analog or a fragment thereof, in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the Batten disease polypeptide, e.g., a humoral response, an antibody response, or a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g., a unique determinant, from a protein represented by SEQ ID NO: 2 or SEQ ID NO: 19.

The invention also includes an antibody preparation, preferably a monoclonal antibody preparation, specifically reactive with an epitope of the Batten disease immunogen or generally of a Batten disease polypeptide.

Also included in the invention is a composition which includes a Batten disease polypeptide (or a nucleic acid which encodes it) and one or more additional components, e.g., a carrier, diluent, or solvent. The additional component can be one which renders the composition useful for *in vitro*, *in vivo*, pharmaceutical, or veterinary use.

In another aspect, the invention provides a substantially pure nucleic acid having, or comprising, a nucleotide sequence which encodes a polypeptide, the amino acid sequence of which includes, or is, the sequence of a Batten disease polypeptide, or analog or fragment thereof.

In preferred embodiments, the nucleic acid encodes a polypeptide having one or more of the following characteristics: at least one biological activity of a Batten disease polypeptide, e.g., a polypeptide specifically reactive with an antibody, or antibody fragment, directed against a Batten disease polypeptide; an amino acid sequence at least 60%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence from SEQ ID NO: 2 or SEQ ID NO: 19; an amino acid sequence more than 85% homologous to an amino acid sequence from SEQ ID NO: 2 or SEQ ID NO: 19; an amino acid sequence essentially the same as an amino acid sequence in SEQ ID NO: 2 or SEQ ID NO: 19, the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; at least 5, preferably at least 10, more preferably at least 20, most preferably at least 50, 100, or 150 contiguous amino acids from SEQ ID NO: 2 or SEQ ID NO: 19; an amino acid sequence which is preferably at least 10, but no more than 100, amino acids in length, and contains one, two, three or more phosphorylation sites; the ability to act as an agonist or an antagonist of a biological activity of a naturally occurring Batten disease polypeptide.

In preferred embodiments: the nucleic acid is or includes the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 18; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO: 18; the nucleic acid is more than 82% homologous with a nucleic acid sequence of

SEQ ID NO:1 or SEQ ID NO: 18; the nucleic acid includes a fragment of SEQ ID NO:1 or SEQ ID NO: 18 which is at least 25, 50, 100, 200, 300, 400, 500, or 1,000 bases in length; the nucleic acid differs from the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 18 due to degeneracy in the genetic code.

In a preferred embodiment the polypeptide encoded by the nucleic acid is a mammalian, e.g., a human or a rodent, e.g., a mouse or a rat, polypeptide.

In a preferred embodiment the polypeptide encoded by the nucleic acid is an agonist which, for example, is capable of enhancing an activity of a naturally-occurring mutant or wild type Batten disease polypeptide. In another preferred embodiment, the encoded polypeptide is an antagonist which, for example, inhibits an undesired activity of a naturally-occurring Batten disease polypeptide (e.g., a polypeptide having an amino acid sequence shown in SEQ ID NO: 2 or SEQ ID NO: 19).

In a preferred embodiment, the encoded Batten disease polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence in SEQ ID NO:2 or SEQ ID NO:19. The differences, however, are such that the encoded Batten disease polypeptide exhibits at least one biological activity of a naturally occurring Batten disease polypeptide (e.g., the Batten disease polypeptide of SEQ ID NO:2 or SEQ ID NO:19).

In preferred embodiments, the nucleic acid encodes a Batten disease polypeptide which includes a Batten disease polypeptide sequence, as described herein, as well as other N-terminal and/or C-terminal amino acid sequences.

In preferred embodiments, the nucleic acid encodes a polypeptide which includes all or a portion of an amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:19, fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence from SEQ ID NO:2 or SEQ ID NO:19.

In preferred embodiments, the encoded polypeptide is a recombinant fusion protein having a first Batten disease polypeptide portion and a second polypeptide portion having an amino acid sequence unrelated to a Batten disease polypeptide. The second polypeptide portion can be, e.g., any of glutathione-S-transferase; a DNA binding domain; or a polymerase activating domain. In preferred embodiments the fusion protein can be used in a two-hybrid assay.

In preferred embodiments, the encoded polypeptide is a fragment or analog of a naturally occurring Batten disease polypeptide which inhibits reactivity with antibodies, or $F(ab')_2$ fragments, specific for a naturally occurring Batten disease polypeptide.

In preferred embodiments, the nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the Batten disease gene sequence, e.g., to render the Batten disease gene sequence suitable for use as an expression vector.

In yet another preferred embodiment, the nucleic acid of the invention hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides from SEQ ID NO:1 or SEQ ID NO:18, or more preferably to at least 20 consecutive nucleotides from SEQ ID NO:1, or more preferably to at least 40 consecutive nucleotides from SEQ ID NO:1 or SEQ ID NO:18.

In a preferred embodiment, the nucleic acid comprises bases 598-814 of SEQ ID NO: 1. Alternatively, the nucleic acid preferable encodes a Batten disease polypeptide comprising amino acid residues 155-226 of SEQ ID NO: 2.

In a preferred embodiment, the nucleic acid encodes a mature polypeptide having a molecular weight of about 48 kDa.

In a preferred embodiment, the nucleic acid encodes a Batten disease polypeptide which includes a leader sequence, e.g., an N-terminal sequence responsible for secretion of the polypeptide from a cell in which it is expressed, or other sequence which is not present in the mature protein. In another preferred embodiment, nucleic acid encodes a Batten disease polypeptide, e.g., the polypeptide of SEQ ID NO: 2 or SEQ ID NO:19, which lacks a leader sequence, e.g., an N-terminal sequence responsible for secretion of the polypeptide from a cell in which it is expressed, or other sequence which is not present in the mature protein.

In another aspect, the invention includes: a vector including a nucleic acid which encodes a Batten disease polypeptide, e.g., a Batten disease polypeptide; a host cell transfected with the vector; and a method of producing a recombinant Batten disease -like polypeptide, e.g., a Batten disease polypeptide; including culturing the cell, e.g., in a cell culture medium, and isolating the Batten disease -like polypeptide, e.g., a Batten disease polypeptide, e.g., from the cell or from the cell culture medium.

In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:18, more preferably having more than 82% homology with a nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:18.

The invention also provides a probe or primer which, e.g., includes or comprises a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence from SEQ ID NO:1 or SEQ ID NO:18, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label group attached thereto. The label group can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 or 20 and preferably less than 20, 30, 50, 100, 150 or 500 nucleotides in length. Preferred primers of the invention include oligonucleotides having a nucleotide sequence shown in any of SEQ ID NOS: 3-15 and 20-58.

In preferred embodiments: the probe or primer is within a deletion, e.g., the 1.02 Kb deletion described herein; the probe or primer is outside a deletion, e.g., the 1.02 Kb deletion described herein; or the probe or primer spans a deletion, e.g., the 1.02 Kb deletion described herein.

In other preferred embodiments, the probe or primer overlaps one of the lesions described herein.

The invention involves nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

In another aspect, the invention features a method of evaluating whether a mammal, for example a primate or a human, is at risk for Batten disease or the misexpression of a Batten disease gene, characterized by, for example, accumulation of autofluorescent lipopigments (ceroid and lipofuscin) in neurons and other cell types leading to progressive loss of vision, seizures and psychomotor disturbances. The method includes detecting, in a tissue of the subject, the presence or absence of a mutation of a Batten disease gene, e.g., a gene encoding a protein represented by SEQ ID NO: 2 ,SEQ ID NO:19, or a homolog thereof. In preferred embodiments: detecting the mutation includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from the gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

For example, detecting the genetic lesion can include: (i) providing a PCR probe, e.g., a radiolabeled PCR probe, amplified from cDNA (e.g., SEQ ID NO: 1 or SEQ ID NO:18) encoding a Batten disease polypeptide and containing a nucleotide sequence which hybridizes to a sense or antisense sequence from the Batten disease gene (e.g., SEQ ID NO: 1 or SEQ ID NO:18), or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the Batten disease gene; (ii) exposing the probe/primer to nucleic acid of the tissue (e.g., genomic DNA) digested with one of many known restriction endonucleases; and (iii) detecting by *in situ* hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion. Alternatively, direct PCR analysis, using primers specific for a Batten disease gene (e.g., a gene comprising the nucleotide sequence shown in SEQ ID NO: 1 or SEQ ID NO:18), can be used to detect the presence or absence of the genetic lesion in genomic DNA from an individual.

In other preferred embodiments, sequencing of the Batten disease gene or fragments thereof can be used to detect lesions described in Table 3 below.

In another aspect, the invention provides a method for detecting in a tissue of a subject, the presence or absence of a lesion, e.g., a deletion, an insertion or a rearrangement, in a Batten disease gene, e.g., a gene encoding a protein represented by SEQ ID NO: 2 ,SEQ

ID NO:19, or a homolog thereof. The method includes: (i) providing a primer which spans the lesion; (ii) amplifying a nucleic acid of the tissue (e.g., genomic DNA) with the lesion spanning primer; and (iii) detecting the presence or absence of the lesion. In preferred embodiments: the deletion is from about 200 to about 2000 bp in size; the deletion is about 1000 bp in size; the deletion has a core haplotype "56" (based on the size of alleles, *D16S299* and *D16S298*, with which it displays close linkage disequilibrium).

In a preferred embodiment, the method further includes either or both of amplifying the nucleic acid of the tissue with a primer located within the lesion, and a second primer located outside the lesion. For example, primers of SEQ ID NOs:20-28 can be used to detect a frequently occuring 1.02 Kb deletion of the Batten disease gene.

In a preferred embodiment, the lesion can be any of lesions described herein, e.g., a 1.02 Kb deletion or those described in Table 3 below.

In another aspect, the invention provides a method of determining if a subject mammal, e.g., a primate, e.g., a human, is at risk for a Batten disease or misexpression of a Batten disease gene, characterized by, for example, accumulation of autofluorescent lipopigments (ceroid and lipofuscin) in neurons and other cell types leading to progressive loss of vision, seizures and psychomotor disturbances. The method includes detecting, in a tissue of the subject, misexpression (e.g., a non-wild type level) of a Batten disease polypeptide or Batten disease polypeptide RNA. In a preferred embodiment, the method utilizes an antibody, such as a monoclonal antibody, specific for a Batten disease polypeptide, or an analog or fragment of a Batten disease polypeptide, to detect misexpression of a Batten disease polypeptide.

In another aspect, the invention features a method of evaluating a compound for the ability to interact with, e.g., bind, a Batten disease polypeptide. The method includes contacting the compound with the Batten disease polypeptide, and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the Batten disease polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with Batten disease polypeptides. It can also be used to find natural or synthetic inhibitors of mutant Batten disease polypeptides.

In brief, a two hybrid assay system (see e.g., Bartel et al. (1993) *Cellular Interaction in Development: A practical Approach*, D.A. Hartley, ed., Oxford University Press, Oxford, pp. 153-179) allows for detection of protein-protein interactions in yeast cells. The known protein, e.g., a Batten disease polypeptide, is often referred to as the "bait" protein. The proteins tested for binding to the bait protein are often referred to as "fish" proteins. The "bait" protein, e.g., a Batten disease polypeptide, is fused to the GAL4 DNA binding domain. Potential "fish" proteins are fused to the GAL4 activating domain. If the

"bait" protein and a "fish" protein interact, the two GAL4 domains are brought into close proximity, thus rendering the host yeast cell capable of surviving a specific growth selection.

In another aspect, the invention features a method of identifying compounds which interact with fragments or analogs of a Batten disease polypeptide. The method includes first identifying compounds which interact with a Batten disease polypeptide, for example, the two hybrid assay described above. These compounds can then be used as "bait" to fish for and identify fragments of the Batten disease polypeptide which also interact, bind, or form a complex with these compounds.

In another aspect, the invention features a method of evaluating an effect of a treatment, e.g., a treatment used to treat a disorder related to the Batten disease gene, e.g., a disorder characterized by progressive loss of vision, seizures and psychomotor disturbances, e.g., Batten disease. The method uses a wild type test cell or organism, or a cell or organism which misexpresses the Batten disease gene or which has a Batten disease transgene, e.g., a transgenic animal. The method includes: administering the treatment to a test cell or organism, e.g., a cultured neural cell, or a mammal, and evaluating the effect of the treatment on a parameter related to an aspect of Batten disease, e.g., a neurodegenerative parameter, such as the accumulation of autofluorescent lipopigments in the cultured neural cell or cells of the mammal, or on the expression of the gene. An effect on the parameter indicates an effect of the treatment.

In another aspect, the invention features a method of making a Batten disease polypeptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring Batten disease polypeptide. The method includes altering the sequence of a Batten disease polypeptide (e.g., SEQ ID NO: 2 or SEQ ID NO:19) by, for example, substitution or deletion of one or more residues of a non-conserved region, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of making a fragment or analog of a Batten disease polypeptide, e.g., a Batten disease polypeptide having at least one biological activity of a naturally occurring Batten disease polypeptide. The method includes altering the sequence, e.g., by substitution or deletion of one or more residues, preferably which are non-conserved residues, of a Batten disease polypeptide, and testing the altered polypeptide for the desired activity.

In another aspect, the invention features a method of treating a mammal, e.g., a human, at risk for Batten disease, e.g., a disorder characterized by neurodegeneration, such as progressive loss of vision, seizures and psychomotor disturbances. The method includes administering to the mammal a therapeutically effective amount of a nucleic acid encoding a Batten disease polypeptide. The nucleic acid can encode an agonist or antagonist of a Batten disease polypeptide.

In another aspect, the invention features a method of treating a mammal, e.g., a human, at risk for Batten disease, e.g., a disorder characterized by neurodegeneration, such as progressive loss of vision, seizures and psychomotor disturbances. The method includes administering to the mammal a therapeutically effective amount of a Batten disease polypeptide. The polypeptide can be an agonist or antagonist of a Batten disease polypeptide.

In another aspect, the invention features, a method of evaluating a compound for the ability to bind a nucleic acid encoding a Batten disease gene regulatory sequence. The method includes: contacting the compound with the nucleic acid; and evaluating ability of the compound to form a complex with the nucleic acid. In preferred embodiments the Batten disease gene regulatory sequence is functionally linked to a heterologous gene, e.g., a reporter gene.

In another aspect, the invention features a human cell, e.g., a neuron, transformed with a nucleic acid which encodes a Batten disease polypeptide.

In another aspect, the invention includes: an expression vector containing a nucleic acid encoding a Batten disease polypeptide (e.g., SEQ ID NO: 2 or SEQ ID NO:19), or an analog or fragment thereof; a cell transformed with an expression vector containing a nucleic acid encoding a Batten disease polypeptide (e.g., SEQ ID NO: 2 or SEQ ID NO:19), or an analog or fragment thereof; and a Batten disease polypeptide made by culturing a cell transformed with an expression vector containing a nucleic acid encoding a Batten disease polypeptide (e.g., SEQ ID NO: 2 or SEQ ID NO:19), or an analog or fragment thereof.

In another aspect, the invention includes a transgenic animal, preferably a mammal, e.g., a mouse, rat, pig or goat, having a Batten disease transgene, e.g., a Batten disease gene having a deletion of all or a part of the wild type Batten disease gene. The transgenic animal can be heterozygous or homozygous for the transgene.

Such a transgenic animal can serve as a model for studying disorders which are related to mutated or mis-expressed Batten disease gene alleles or for use in drug screening. For example, the invention includes a method of evaluating the effect of the expression or misexpression of a Batten disease gene on a parameter related to Batten disease. The method includes: providing a transgenic animal having a Batten disease transgene, or which otherwise misexpresses a Batten disease gene; contacting the animal with an agent; and evaluating the effect of the transgene on the parameter related to Batten disease polypeptide metabolism.

A "heterologous promoter", as used herein is a promoter which is not naturally associated with the Batten disease gene.

A "purified preparation" or a "substantially pure preparation" of a Batten disease polypeptide, or a fragment or analog thereof, as used herein, means a Batten disease polypeptide. or a fragment or analog thereof, that has been separated from on or more other proteins, lipids, and nucleic acids with which the Batten disease polypeptide naturally occurs.

Preferably, the polypeptide, or a fragment or analog thereof, is also separated from substances which are used to purify it, e.g., antibodies or gel matrix, such as polyacrylamide. Preferably, the polypeptide, or a fragment or analog thereof, constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10. or 100 μg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

A "treatment", as used herein, includes any therapeutic treatment, e.g., the administration of a therapeutic agent or substance, e.g., a drug.

The "metabolism of a substance", as used herein, means any aspect of the, expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the structure or distribution of other substances.

A "substantially pure nucleic acid", e.g., a substantially pure DNA encoding a Batten disease polypeptide, is a nucleic acid which is one or both of: not immediately contiguous with one or both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid sequence with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional Batten disease sequences.

"Homologous", as used herein, refers to the sequence similarity between two polypeptide molecules or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine; then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10, of the positions

in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

The terms "peptides", "proteins", and "polypeptides" are used interchangeably herein.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., one or more Batten disease polypeptides), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of the selected nucleic acid, all operably linked to the selected nucleic acid, and may include an enhancer sequence.

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgene. The transgene can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence, such as the Batten disease gene, operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as neurons. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

"Unrelated to a Batten disease amino acid or nucleic acid sequence" means having less than 30% homology, less than 20% homology, or, preferably, less than 10% homology with a Batten disease sequence disclosed herein.

A polypeptide has "at least one biological activity of a Batten disease polypeptide" if it has one or more of the following properties: (1) the ability to react with an antibody, or antibody fragment, specific for (a) a wild type Batten disease polypeptide, (b) a naturally-occurring mutant Batten disease polypeptide, or (c) a fragment of either (a) or (b); (2) the ability to prevent, treat or correct a disorder associated with Batten disease, including,

for example, neurodegenerative disorders characterized by progressive loss of vision, seizures and psychomotor disturbances; or (3) the ability to act as an antagonist or agonist of the activities recited in (1) or (2).

"Misexpression", as used herein, refers to a non-wild type pattern of Batten disease gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing, size, amino acid sequence, post-transitional modification, stability, or biological activity of the expressed Batten disease polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the Batten disease gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As described herein, one aspect of the invention features a pure (or recombinant) nucleic acid which includes a nucleotide sequence encoding a Batten disease polypeptide, and/or equivalents of such nucleic acids. The term "nucleic acid", as used herein, can include fragments and equivalents. The term "equivalent" refers to nucleotide sequences encoding functionally equivalent polypeptides or functionally equivalent polypeptides which, for example, retain the ability to react with an antibody specific for a Batten disease polypeptide. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants, and will, therefore, include sequences that differ from the nucleotide sequence of Batten disease shown in SEQ ID NO: 1 due to the degeneracy of the genetic code.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are described in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc.. N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and

M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

The Batten disease gene and polypeptide of the present invention are useful for studying, diagnosing and/or treating Batten disease. For example, the gene (or fragment thereof) can be used to detect and study genetic mutations or gene transcripts commonly associated with Batten disease, as described in detail below. The gene (or fragment thereof) can be used in gene replacement therapy to correct the absence of a wild type Batten disease gene (e.g., to reconstitute the function of, enhance the function of, or alternatively, antagonize the function of a Batten disease polypeptide in a cell in which the polypeptide is misexpressed). The gene (or fragment thereof) can be used to prepare antisense constructs capable of inhibiting expression of a mutant or wild type Batten disease gene encoding a polypeptide having an undesirable function. Alternatively, a Batten disease polypeptide can be used to raise antibodies capable of detecting proteins or protein levels associated with Batten disease. A Batten disease polypeptide can be administered to a patient afflicted with Batten disease to correct the absence of a wild type Batten disease polypeptide, or as an agonist to enhance the activity of a wild type Batten disease polypeptide.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Figures

Figure 1 is a schematic representation of the *CLN3* candidate region on chromosome 16p12.1. The positions of selected DNA microsatellites used for linkage and haplotype analysis are indicated. Individual cosmids (NL11A, NL60D3) of cosmid contig CNL/343.1, which contains *D16S298* and *D16S48*, and cosmid contig C182, which contains *D16S299*, are indicated by horizontal lines. Three YACs (Cy21B11, Cy302G12, Cy85D3) that form part of a 980 kb contig spanning the candidate region are also indicated by horizontal lines.

Figure 2 is a restriction map of cosmid NL11A. The genomic extent of cDNA2-3 is shown below the map (arrow indicating the direction of transcription). The position of the 3.12 STS, the microsatellite marker *D16S298*, and the overlapping cosmid NL60D3 are shown above the restriction map.

Figure 3 is the nucleotide sequence of cDNA2-3. The predicted protein is shown below the DNA sequence, assuming that translation begins at the first in-frame

methionine of the long open reading frame. Four potential N-linked glycosylation sites are indicated by a dashed line at residues 49, 71, 85, and 310. Two potential glycosaminoglycan sites are indicated by the dotted lines at residues 162 and 186. Potential N-myristoylation sites are indicated by(#). Serine and threonine residues that are potentially phosphorylated by cAMP- and cGMP-dependent protein kinases (%), or protein kinase C (*), or casein kinase 2 (^) are indicated. The polyadenylation site at base 1666 is indicated by the $. cDNA sequence deleted in the "56" deletion (bases 598-814) is boxed.

Figure 4 is a Mendelian inheritance diagram showing segregation of the "56" haplotype (deletion) in a two-generation Batten Disease family.

Figure 5 is a diagram showing the 1.02 kb genomic deletion in disease chromosomes bearing the "56" haplotype. The sequences bordering the deletion are shown. The deletion covers two exons and flanking intronic sequence and leads to the deletion of 217 bp of coding sequence. The two flanking exons are spliced together to read CCTGTGTGCTATTTC (SEQ ID NO: 17) in the patient mRNA. Position of primers used to delineate the deletion are also indicated. Hatched boxes represent exons. The boxes indicate the positions of Alu-Sx sequences. The deletion breakpoints are shown by the arrows, and deleted sequences are shown in italics.

Figure 6 is a schematic representation of the genomic deletions of the 2-3 gene. Position of primers used to delineate the deletions are indicated.

Figure 7 is a schematic representation of a direct detection of the major deletion of the CLN3 gene. Normal and deletion alleles of CLN3. Primer 2.3LR3 is located within the deleted region whereas primer CLN3mut756R is spanning the deletion junction. The allele-specific PCR products are indicated.

Figure 8 is a schematic representation of the location of mutations in CLN3. The mutations are shown in relation to their position in the exons of the cDNA. Those above the cDNA are point mutations in the ORF, those below deletions, insertions or point mutations in introns. Those in bold are missense mutations. Those in italics are mutations in introns. Three are large genomic deletions, the deleted nucleotides shown relate to the cDNA only.

Figure 9 is a schematic representation of the predicted structure of CLN3 protein. The location of the six missense mutations is shown.

Figure 10 is a cromatograph depicting direct sequence analysis of exon 7 in an unaffected control (lower panel) and patient L29 (upper panel). The * indicates the point mutation (C619G).

Detailed Description

The invention provides the sequence of a gene responsible for Batten disease, hereafter referred to as *CLN3*, or as the Batten disease gene. The *CLN3* gene possesses an open reading frame of 1314 bp (SEQ ID NO: 1) encoding a polypeptide having a predicted length of 438 amino acids (SEQ ID NO: 2) and a predicted molecular weight of about 48 kDa (mature protein), with no significant similarity to previously described proteins.

The gene is disrupted by a small (1.02 kb) deletion on all Batten disease chromosomes with a core haplotype "56" (based on the size of alleles, *D16S299* and *D16S298*, with which it displays close linkage disequilibrium), and by independent deletion in the Moroccan patient described below.

Isolation and characterization of Batten Disease cDNA

To clone a cDNA corresponding to the Batten disease gene (*CLN3*), a cosmid (NL11A) which encompasses the *D16S298* allele (known to be closely linked to *CLN3*) was targeted. Exon amplification was used to isolate a 180 bp exon from NL11A. This exon was then used to screen a fetal brain cDNA library (Stratagene), yielding a 1.7 kb cDNA clone (cDNA2-3) (SEQ ID NO: 1).

Southern blot and PCR analyses of genomic and cosmid DNAs confirmed that the 1.7 kb cDNA (SEQ ID NO: 1) was contained in NL11A (Figure 1). As shown in Figure 2, a PCR product corresponding to the 3' end of the cDNA hybridized to a 2.8 kb PstI fragment, while a PCR product corresponding to the 5' end of the cDNA hybridized to a 1.95 kb PstI fragment. This indicated that the 1.7 kb cDNA was contained within NL11A and that transcription proceeded toward *D16S299*. PCR amplification of individual PstI fragments of NL11A, using both *D16S298* microsatellite primers and primers for the adjacent 3.12 STS, placed *D16S298* on a 1.3 kb PstI fragment previously shown to be contained within the deletion of a Moroccan patient affected with Batten disease. This fragment was not detected by cDNA2-3 (SEQ ID NO: 1), but consisted of intron sequences mapping between bases 1193 and 1194 of the cDNA (SEQ ID NO: 1) (Figure 3). Thus, cDNA2-3 (SEQ ID NO: 1) was found to span the *D16S298* locus and to overlap with the deletion found in the Moroccan patient.

Northern blot analysis using cDNA2-3 as a probe revealed a 1.7 kb transcript in polyA-mRNA isolated from a wide variety of human tissues including heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. This result was consistent with the cDNA clone likely being full-length. The transcript was not detected in cultured

lymphoblasts and fibroblasts by Northern blot analysis, but was detectable by RT-PCR analysis of polyA-mRNA isolated from such cell lines. A "zoo" blot containing genomic DNAs from several animal species showed that this gene is conserved in mammals. Strong signals were obtained from mouse, sheep, dog, cow, and pig.

Sequence Analysis of Batten Disease cDNA

Figure 3 shows the nucleotide sequence of cDNA2-3 (SEQ ID NO: 1) which contains 1,689 base pairs (bp) and has a 47 base polyA tail. The cDNA clone has a predicted open reading frame of 1314 bp begins with a potential initiator ATG codon at base 138 and ends with a TGA termination codon at base 1452. An in-frame stop codon is located 36 bases upstream of the initiator site and a consensus polyadenylation site is located at base 1666. The predicted product of the cDNA is a protein of 438 amino acids (SEQ ID NO: 2) with a molecular weight of about 48 kDa. Table 1 lists the sequences and locations of PCR primers derived from this cDNA sequence and used in the studies described below.

TABLE 1

Primer	Location in cDNA	Sequence 5' -> 3'
Forward:		
P1 (SEQ ID NO: 3)	39	TTGATCCTTGTCACCTGTCG
F2 (SEQ ID NO: 4)	552	TTCGTCCTGGTTGCCTTT
F4 (SEQ ID NO: 5)	676	TGATCTCCTGGTGGTCCTCA
F5 (SEQ ID NO: 6)	778	TGTCCATGCTGGGTATCCCT
P2 (SEQ ID NO: 7)	860	GAAGAAGAAGCAGAGAGCGC
F9 (SEQ ID NO: 8)	888	CAGCCCCTCATAAGAACCGA
GF1 (SEQ ID NO: 9)	1470	GGACGCAGGTCACATTCA
Reverse:		
R1 (SEQ ID NO: 10)	656	AGTGAGGGAGAGGAAGGTGA
P3 (SEQ ID NO: 11)	880	CGCTCTCTGCTTCTTCTTCC
R5 (SEQ ID NO: 12)	1246	CTTGGCAGAAAGCCGAAC
R3 (SEQ ID NO: 13)	1612	CCCCTGCAAGGAAACAAG
GR1 (SEQ ID NO: 14)	1661	GGCATGATGCCAGGAAAGA
P5 (SEQ ID NO: 15)	1669	ATTCAGAAGGCATGATGCC

The Batten disease cDNA sequence (SEQ ID NO: 1) was compared against GenBank and dbEST databases using BLASTN (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410) and FASTA (Pearson et al. (1988) *Proc. Natl. Acad. Sci. usa* 85:2444-2448) sequence alignment algorithms. These searches revealed no significant similarities to genes of known function. However, near identity (>95% similarity) was found to 13 ESTs (F11432, F12401,

T74504, T08995, R12998, Z42735, T47968, D20292, T47969, T97772, F09095, F10019, T61330) isolated from 5 independent cDNA libraries (infant brain, fetal spleen, fetal liver and spleen, adult liver, and promyelocyte cell line HL60). Three pairs of ESTs (F11432/F09095, F12401/F10019, and T47968/T47969) are 5' and 3' sequences of three cDNA clones. Five ESTs (F11432, F12401, T74504, T08995, and R12998), all isolated from a normalized infant brain cDNA library (Soares et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:9228-9232) are missing bases 184-262 of cDNA2-3 (SEQ ID NO: 1) (Figure 3). If the same initiator ATG is used, this transcript is expected to produce a truncated protein of only 27 amino acids. Thus, it is unlikely to be the result of normal RNA splicing. The physiological significance of this variant is unclear, since its relative abundance may be exaggerated by preparation of the normalized cDNA library.

The predicted protein sequence (SEQ ID NO: 2) of the polypeptide encoded by cDNA2-3 (SEQ ID NO: 1) was compared against the Swiss-Prot database using BLASTP and Smith-Waterman (Smith et al. (1981) *J. Mol. Biol.* 147:195-197) sequence alignment algorithms and against the predicted translation products of GenBank database using TBLASTN. In all these cases, no significant similarities were found to known proteins. A search of the BLOCKS database (version 8.0; Henikoff et al. (1994) *Genomics* 19:97-107) for motifs found only single blocks of homology for any group of proteins and this could be attributed to chance. A search for protein motifs in the *CLN3* protein using the ProSite Database (version 12.2) revealed pattern matches for 4 N-glycosylation sites, 2 glycosaminoglycan attachment sites, 2 cAMP- and cGMP-dependent protein kinase phosphorylation sites, 6 protein kinase C phosphorylation sites, 8 casein kinase II phosphorylation sites, and 12 N-myristoylation sites (Figure 3). Hydropathy calculations (Kyte et al. (1982) *J. Mol. Biol.* 157:105-132) predicted 5 hydrophobic regions which may be potential membrane spanning regions at amino acids 38-61, 93-233, 278-310, 345-399, and 408-438 of the encoded polypeptide (SEQ ID NO: 2).

The Common Mutation in the Batten Disease Gene is a Small Deletion

To screen for possible deletions, insertions, and other chromosomal rearrangements associated with *CLN3*, conventional Southern blots of restriction-digested DNA from unrelated Batten disease patients were scanned. A panel of PstI-digested patient DNAs were hybridized with PCR probes P1-P3 (SEQ ID NOS: 3 and 11) and P2-P5 (SEQ ID NOS: 7 and 15) (Table 1) representing the 5' and 3' halves of the cDNA (SEQ ID NO: 1), respectively. When the P1-P3 fragment was used as probe, affected individuals homozygous for the "56" *D16S299/D16S298* haplotype displayed the loss of a 3.8 kb PstI fragment and the gain of a novel 2.8 kb fragment. When the P2-P5 fragment was used as probe, no difference was detected between controls and the homozygous "56" haplotype affected. Analysis of 148 control chromosomes, including 7 with the "56" haplotype, revealed no alterations. The

affected individuals bearing a "56" chromosome also displayed altered fragments with HindIII and PvuII digestion, suggestive of a small (~1000 bp) genomic deletion of the "56" chromosome. Figure 4 illustrates the Mendelian inheritance of this deletion in a two-generation Batten Disease family segregating the "56" haplotype. The chromosomes segregating in this pedigree have been distinguished by extensive typing with polymorphic markers in 16p12.1-11.2.

To determine the effect of this genomic deletion on the cDNA2-3 transcript, we performed PCR amplification of RT-cDNA from patients homozygous for the "56" haplotype as follows: RT-cDNA was prepared from cytoplasmic RNA isolated from the peripheral blood lymphocytes of 6 normal controls, the fibroblasts of 1 normal control, and the fibroblasts from 4 patients homozygous for the "56" haplotype. PCR products were fractionated on 1-1.5% gels and transferred to Hybond N+ (Amersham) membranes. Blots were hybridized with the radiolabeled PCR fragments amplified from the cDNA2-3 clone.

Patients homozygous for the "56" haplotype yielded an P1-P3 RT-PCR product ~200 bp smaller than the corresponding RT-PCR product from control individuals. In control individuals, amplification with either P1-P3 (SEQ ID NOS: 3 and 11) or P2-P5 (SEQ ID NOS: 7 and 15) primer set yields a ~800 product, although these fragments contain different sequences. Thus, the P1-P3 primer pair (SEQ ID NOS: 3 and 11) yielded a novel product, ~ 200 bp smaller than that predicted from the cDNA sequence and found in all non-"56" normal controls, and RT-PCR amplification with the P2-P5 primer pair yielded identical ~800 bp products in affected and controls.

DNA sequence analysis of the P1-P3 product from 5 homozygous "56" patients showed in all cases a 217 bp deletion, from base 598 to base 814 (SEQ ID NO: 16) of the cDNA (SEQ ID NO: 1) (Figure 3). The DNA sequence of the RT-cDNA from 4 control individuals revealed no evidence of deletion, matching the cDNA2-3 sequence. Deletion of these 217 bases of coding sequence (SEQ ID NO: 16) produces a frameshift, generating a TAA termination codon 84 bp downstream of the deletion junction. The predicted translation product is a truncated protein of 181 amino acids consisting of the first 153 residues of the protein followed by 28 novel amino acids before the stop codon.

DNA sequence analysis of the genomic fragment containing this deletion from a "56" homozygous patient revealed the loss of 1.02 kb of genomic sequence (Figure 5). The intron sequence immediately 5' to the deletion is 91% homologous to bases 84-290 of the Alu-Sx family consensus sequence in the 5' to 3' orientation. The A-rich sequence at the 3' end of this Alu sequence includes a GA_4 repeat sequence within the 5' deleted segment. The sequence at the 3' portion of the deleted region is 87% homologous to bases 1-290 of the Alu-Sx sequence also in the 5' to 3' orientation and contains a GA_4 repeat sequence within the A-rich sequence of the 3' tail. Included in this deletion are 217 bp of the open reading frame (bp 598-815 (SEQ ID NO: 16) (Figure 3), corresponding to two exons.

Screening for the "56" Deletion in the Batten Disease Gene

PCR amplification of genomic DNA with primers F2 (SEQ ID NO: 4) and P3 (SEQ ID NO: 11) flanking the cDNA deletion of "56" patients (described above) produced a 3.5 kb product from normal chromosomes and a 2.5 kb product from the chromosomes with the 1.02 kb deletion described above. The presence of the 1.02 kb deletion associated with this "56" *D16S299/D16S298* haplotype was tested for in 81 unrelated Batten patients representing 24 haplotypes and originating from 16 countries, as shown below in Table 2. Forty-six were homozygous for the "56" haplotype, 24 were heterozygous for the "56" haplotype, and 11 did not carry the "56" haplotype on either chromosome. In all 70 patients with a "56" affected chromosome, the 2.5 kb fragment was detected, and in all 46 homozygotes for this haplotype, no normal size product was produced. Smaller numbers of chromosomes bearing closely-related haplotypes (66, 36, 46, 57, and 55) also carried this deletion, suggesting that these chromosomes most probably derived from the "56" haplotype by mutation of the polymorphic marker or recombination. Additional affected chromosomes bearing the "66" and "46" haplotypes apparently possess mutations independent of the "56" chromosomes, as they do not carry this deletion. Thus, the 1.02 kb genomic deletion of the *CLN3* gene associated with the "56" haplotype is the most common mutation in Batten disease, accounting for 81% of disease chromosomes tested to date.

TABLE 2

HAPLOTYPE CHROMOSOMES *D16S299/D16S298*	PCR AMPLIFICATION PRODUCT (KB) 2.5	3.5	NO. OF
56	+	-	116
	-	+	0
66	+	-	4
	-	+	7
36	+	-	4
	-	+	0
46	+	-	2
	-	+	1
65	+	-	1
	-	+	0
67	+	-	1
	-	+	0
57	+	-	2
	-	+	0
55	+	-	1
	-	+	0
Other haplotypes	+	-	1
	-	+	22
Total No. Chrs.			162

Genomic PCR was carried out using primer pair F2-P3 (SEQ ID NOS: 4 and 11) at bases 553 and 880, respectively, of the cDNA2-3 sequence (SEQ ID NO: 1; Fig. 3). PCR amplification was carried out as described below in the Experimental Methods.

Other Mutations Disrupting the Batten Disease Gene

Haplotype analysis of Finnish patient L199Pa revealed one "56" chromosome and one "6null" chromosome exhibiting absence of any D16S298 allele (see Experimental Procedures for clinical details). Southern blot analysis of this patient revealed two alterations: the 1.02 kb deletion typical of the "56" chromosomes and a second deletion present on the chromosome missing D16S298 that results in the formation of a novel 1.5 kb junction fragment. This junction fragment combines sequences from an upstream 1.1 kb PstI

fragment detected by the cDNA probe and from a PstI fragment 3' to D16S298 that contains only intron sequence. PCR analysis of patient DNA using the intron primer intR14 (5'-aggaaggaggctggaggata-3')(SEQ ID NO:58) and cDNA primer F9 (SEQ ID NO:8) confirmed an ~3 kb deletion, including the entire 1.3 kb PstI fragment containing D16S298. RT-cDNA from this second mutant allele was selectively amplified using primer R5 (SEQ ID NO:12) and primer F5 (SEQ ID NO:6) which is deleted on the "56" chromosome. The amplified product revealed the absence of 266 bp of coding sequence between bases 928-1193 of the cDNA, generating a TGA termination codon 84 bp downstream of the deletion junction. The predicted translation product is a truncated protein of 291 amino acids consisting of the first 263 amino acids of the protein followed by 28 novel amino acids before the stop codon. Partial DNA sequence analysis of the genomic fragment containing this ~3 kb deletion has confirmed the loss of bases 928-1193 of the cDNA. The sequences bordering this deletion have not yet been defined.

A homozygous deletion of the *D16S298* locus in a Batten patient of Moroccan origin (NCL39.3) was previously described by Taschner et al. (1995(*Am. J. Med. Genet.* 57:333-337. Although the size of the deletion was not determined, it did include the 1.3 kb PstI fragment containing *D16S298* that has proved to be within an intron of the *CLN3* candidate gene. PCR amplification of genomic DNA with primers F2 (SEQ ID NO: 4) and R3 (SEQ ID NO: 13) yielded a 1.1 kb fragment instead of the expected ~ 7 kb fragment. Additional PCR amplifications using nested primers on either the 5' (F4-R3) (SEQ ID NOS: 5 and 13) or 3' (GF1-GR1) (SEQ ID NOS: 9 and 14) sides gave no product, suggesting a deletion in the Moroccan patient of about 6 kb which starts between F2 (SEQ ID NO: 4) and F4 (SEQ ID NO: 5) and ends between GF1 (SEQ ID NO: 9) and R3 (SEQ ID NO: 13). The locations of the two deletions described in these studies and the PCR primers used to analyze them are summarized in Figure 6.

Single stranded conformation polymorphism (SSCP) was performed to scan the CLN3 gene for further mutations. Patient L198Pa (see Experimental Procedures for clinical details) is heterozygous with one "56" chromosome and one "76" (D16S299/D16S298) chromosome. This patient exhibited a mobility shift in a 73 bp exon corresponding to bases 598 - 670 of the cDNA. This exon is one of those deleted on the "56" chromosome. Nucleotide sequence analysis showed a G-> C transition at +1 of the splice donor site following the exon. Analysis of the parents of patient L198Pa showed the father (haplotype 76/46) to be a heterozygous carrier of this mutation. Transcriptional analysis is pending the availability of blood samples from this family.

Analysis of the Batten Disease Gene

The data described above demonstrates that the Batten disease gene mutation associated with *D16S299/D16S298* "56" haplotype is a 1.02 kb deletion that implicates

cDNA2-3 as the product of *CLN3*. This deletion involves the 3' end of two Alu-Sx elements and the following GA4 sequence and may therefore have arisen by recombination involving bordering Alu sequences, a mechanism for which other examples exist in human disease (e.g., Rudiger et al. (1995) *Nucleic Acids Res.* 23:256-60). The deletion mutation is found on all "56" affected chromosomes examined to date, and on several chromosomes with related haplotypes, accounting for 81% of Batten disease chromosomes.

With the notable exceptions of patient NCL39.3 (Moroccan), Southern blot and long-range PCR analyses of patients with chromosomes lacking the 1.02 kb deletion have failed to reveal additional genomic rearrangements. These results suggest that these affected chromosomes most likely carry point mutations, small deletions, or regulatory mutations of *CLN3*. The independent deletions in NCL39.3, which encompasses the *D16S298* microsatellite locus, provide the strongest confirmatory evidence that cDNA2-3 is the product of *CLN3*.

Homology have been found at the nucleotide or amino acid level with mouse. dog, *S. cerevisiae* and *C. elegans* genes. Diverse approaches may now be used to explore the Batten disease polypeptide's normal physiological role. For example, the conservation of coding sequences across species should allow the identification of homologous sequences and target conserved domains of functional significance.

The presence of several potential phosphorylation sites suggests that the protein may undergo phosphorylation as a prerequisite for binding additional protein(s). The PSORT program (version 6.3; Nakai et al. (1992) *Genomics* 14:897-911) for prediction of protein localization sites indicates that the *CLN3* protein may be a membrane spanning protein having 6 transmembrane segments (Heijne et al. (1988) *Euro. J. Biochem.* 174:671-678), a possibility supported by hydropathy calculations that suggest the presence of several hydrophobic domains and by numerous potential N-glycosylation and N-myristoylation site.

The deletions identified to date are predicted to remove over 100 amino acids from the C-terminal portion of the Batten disease polypeptide, suggesting that its normal function would be severely compromised in the disease. However, it is also conceivable that the disease phenotype may involve abnormal accumulation of truncated Batten disease polypeptide products rather than, or in addition to, direct loss of protein function. The *CLN3* gene is expressed not only in the brain, the site of massive neuronal cell death in Batten patients, but also in a wide range of tissues. Consistent with this, inclusion bodies have been found in many Batten disease tissues in addition to the brain. In addition, Palmer et al (1992) *Am. J. Med. Genet.* 42:561-567 demonstrated the abnormal accumulation of subunit 9 of mitochondrial ATPase in these inclusions. However, experiments mapping the subunit 9 genes P1 and P2 to chromosomes 17 and 12, respectively, (Dyer et al. (1993) *Biochem J.* 293:51-64) and P3 to chromosome 2 (Yan et al (1994) *Genomics* 24:375-377) excluded these genes as the site of the Batten disease defect. It will now be of interest to determine whether

the Batten disease polypeptide encoded by *CLN3*, or fragments thereof, also accumulate in the disorder. Similarly, various biochemical approaches have suggested that Batten disease involves perturbations in several metabolic pathways including, for example, lipid peroxidation (Siakotos et al. (1988) *Am. J. Med. Genet Suppl.* 5:171-181), metabolism of dolichol-linked oligosaccharides (Hall et al. (1985) *J. Inherited Metab. Dis.* 8:178-183), and lysosomal proteinase activity (Wolfe et al. (1987) *Chem. SCr* 27:79-84). Whether these diverse biochemical phenotypes are the result of the primary gene defect or are secondary effects of the disease process can now be examined as a result of the present invention.

Because of the slow progression of symptoms in Batten disease and its similarity to other NCL subtypes and neurologic disorders, diagnosis is often missed or delayed. Current diagnostic protocols call for examination of skin biopsies for hallmark fingerprint profiles in inclusion bodies, a technically demanding procedure. Since the demonstration of linkage disequilibrium, carrier detection by haplotype analysis has been possible. The direct PCR assay for the "56" Batten disease deletion, described above, will improve the reliability of the diagnosis for the majority of Batten disease patients and provide families with the opportunity for pre-natal and carrier testing.

The identification and isolation of the Batten disease gene provided by the present invention is the first step toward understanding the pathology underlying this complex disorder. The cDNA clone, cDNA2-3, will provide the basis for analyzing the role of the *CLN3* polypeptide in both normal and disease cells and a starting point for the design of rational therapies. Moreover, the availability of cDNA2-3 will allow the study of Batten disease polypeptides encoded by *CLN3*, and may reveal the underlying cause of the other ceroid lipofuscinoses and provide new insights into the mechanisms involved in other neurodegenerative disorders.

Isolation and Chromosomal Mapping of a Mouse Homolog of the CLN3 gene

In order to create a mouse model of Batten disease, a mouse homolog of the human *CLN3* gene was cloned and mapped.

A murine teratocarcinoma cDNA library (Stratagene) was screened by plaque hybridization with the human Batten disease cDNA clone 2-3 as probe, yielding a 1639-bp cDNA, clone mtc7 (SEQ ID NO:18). Clone mtc7 was sequenced manually by the dideoxy method on both strands. The DNA sequence analysis revealed 82% identity between the mouse (SEQ ID NO:18) and the human cDNA coding sequences (SEQ ID NO:1). Like its human homolog, clone mtc7 contains a predicted open reading frame (ORF) of 1314 bp, beginning with a potential initiator ATG codon at base 142 and ending with a TGA termination codon at base 1456. An in-frame stop codon is located 54 bases upstream of the initiator ATG. The cDNA has a consensus polyadenylation site (AATAAA) located at bases 1617-1622 and a 19-base poly(A) tail. The ORF encodes a predicted protein product of 438

amino acids (SEQ ID NO:19) with a high degree of similarity (85% identity) to the human CLN3 protein (SEQ ID NO:2). The four potential N-glycosylation sites found in the human sequence are conserved in the mouse at amino acid residues 49-52 (NFSY), 71-74 (NQSH), 85-88 (NSSS), and 310-313 (NTSL).

mtc7 cDNA was used as a probe to map *CLN3* genetically in the mouse. The map location of *Cln3* was determined by segregation analysis of a mouse interspecific backcross DNA mapping panel derived from matings of (C57BL/6J x SPRET/Ei) F1 females with SPRET/Ei males and designated MMR-BSS. The MMR-BSS panel consists of 144 individuals that have been typed for more than 300 different polymorphic loci (Johnson et al., *Mamm. Genome* 5:670-687, 1994). Probe labeling, blotting, and hybridization conditions used in the present study were the same as previously described (Johnson et al., *Genomics* 12:503-509, 1992). Southern blot analyses using the mouse cDNA probe detected polymorphic, strain-specific PstI restriction fragments. In C57BL/6J DNA, fragment sizes were 4.8, 3.1, 2.5, 1.6, and 1.0 kb; in SPRET/Ei DNA they were 6.8, 3.1, 2.2, and 1.0 kb. The presence or absence of the C57BL/6J-specific 4.8-kb fragment was used to assign *Cln3* genotypes of backcross progeny. Genetic linkage was analyzed by comparing the segregation pattern of *Cln3* genotypes among the backcross progeny with those of previously mapped loci. The computer program Map Manager (Manly, K.F., *Mamm. Genome* 4:303-313, 1993) was used to perform linkage and haplotype analysis. Gene order on a chromosome was determined by minimizing the number of double crossover events.

Linkage of *Cln3* was found with markers on mouse Chromosome 7. *Cln3* mapped about 16 cM distal to Tyr (tyrosinase) between *D7Mit9* and *D7Mit43*. According to the mouse Chromosome 7 Committee report (Brilliant et al., *Mamm. Genome* 5:S104-S123, 1994), this position places *Cln3* about 60 cM distal to the Chr 7 centromere in a region containing genes whose homologs map to human chromosome 16cp12, where the human Batten disease gene, *CLN3*, has been mapped . The results of low-stringency genomic Southern blot analysis are consistent with the presence of only one gene in the mouse that is closely related to the human Batten disease cDNA.

It has been suggested that the motor neuron degeneration (Mnd) mutation in the mouse may be a model for Batten disease (Bronson et al., *Ann. Neurol.* 33:381-385, 1993). Mice homozygous for the Mnd mutation become blind by 2 months of age, develop spastic paresis and paralysis by 1 year, and exhibit the abnormal accumulation of subunit c in sudanophilic storage bodies. The Mnd mutation has been mapped to mouse Chromosome 8 (Messer et al., *Genomics* 18:797-802, 1992). On the basis of the mapping results presented herein, it has been concluded that Mnd and Cln3 are unique loci.

The degree of identity between the human and mouse CLN3 coding sequences indicates that the protein most likely serves the same function in the mouse as in humans. Isolation and characterization of the mouse *Cln3* gene will allow for construction of vectors

for targeted disruption by homologous recombination in embryonic stem cells. Generation of - / - mice should allow for study of the detailed pathogenesis of Batten disease.

Diagnosis of Batten Disease

The major Battens disease mutation is a 1 kb deletion, which is found in 81% of affected chromosomes. Direct gene analysis with PCR primers which flank the deletion can be used for prenatal diagnosis (Munroe et al., *Lancet* 347: 1014-15, 1996) This often results in preferential amplification of the deletion allele compared to the normal due to the large difference in size between the products and may give false positive results.

Therefore, an allele-specific PCR test which allows the simultaneous detection of normal and major deletion alleles of CLN3 was designed. The test uses one primer spanning the deletion junction in combination with a second primer within the deletion and a third primer outside the deletion to follow the segregation of the major deletion within the family of a Batten's disease patient (Fig. 7).

PCR analysis was carried out on 50 ng genomic DNA in a total volume of 25 µl at a final concentration of 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each dNTP, 0.004 U/µl of SuperTaq (HT Biotechnology Ltd., Cambridge, UK), in the presence of 5 pmol of primers 2.3LR3 (5'-GGGGGAGGACAAGCACTG-3'(SEQ ID NO:20)) and 2.3IntF7 (5'-CATTCTGTCACCCTTAGAAGCC-3'(SEQ ID NO:21)) and 4 pmol of primer CLN3mut756R (5'GGACTTGAAGGACGGAGTCT-3'(SEQ ID NO:22)). Denaturation was 3 min at 94°C, annealing for 2 min at 56°C, and extension for 1 min at 72°C, with a final extension for 10 min. The following primers can also be used in the allele-specific PCR test: IntF6 (5'-GGAGCCTCTATGAGCTGATACTG-3'(SEQ ID NO:23)), 6905F (5'-TTCGTCCTGGTTGCCTTT-3'(SEQ ID NO:24)); 6334R (5'-CCTGATGAGATGCTAGCGAA-3'(SEQ ID NO:25)), CLN3mut756F (5'-AGACTCCGTCCTTTCAAGTCC-3'(SEQ ID NO:26)), and IntR7 (5'-TTACACATTCGAGGCCAACCT-3'(SEQ ID NO:27)).

The allele-specific PCR test allows early confirmation of the clinical diagnosis in the majority of the Batten patients which is important for correct prognosis and genetic counseling, and may help to prevent the birth of additional patients. In addition, this test can be used to detect carriers of the major deletion in the general population which is important for unrelated partners of proven carriers.

Experimental Procedures

Patients and Cell Lines

Patients with Batten Disease were identified through contacts with volunteer parents' organizations and through clinical referrals. Diagnoses were confirmed using standard criteria (Boustany et al (1988) *Am. J. Med. Genet. Suppl.* 5:47-58; Santavuori

(1988) *Brain Dev.* 10:80-83). The establishment of lymphoblastoid cell lines was previously described (Anderson et al. (1984) *In Vitro* 20:856-858). The Finnish patient L199Pa had a normal birth and early childhood. At age 6.5, he was referred to the University of Helsinki Clinic and Children's Hospital (Dr. Pirkko Santavuori) because of failing vision. Electroretinogram was abolished and the visual evoked potential (VEP) abnormal with delayed latency. Slight motor clumsiness and muscular hypotonia were found. Vacuolated lymphocytes were positive on repeated examinations. From age 11, he had generalized epileptic seizures that were well controlled by sodium valproate-clonazepam. At age 16 MRI showed slight central, cortical, and cerebellar atrophy. The patient is still able to walk independently, but jumping has become difficult. He has finished school and is working in a day care center.

The Finnish patient L198Pa had an uneventful birth and early childhood. Since the age of 7, she has experienced progressive visual failure. At age 9, she showed abnormal MRI. Vacuolated lymphocytes were repeatedly observed and electronmicroscopy of a rectal biopsy specimen showed inclusions typical for Batten Disease. She has been on sodium valproate medication since the age of 9, when she experienced her only seizure. Recent examination at the age of 13 showed that her motor status is good but that her mental decline has been relatively fast.

The Moroccan patient has been previously described (Taschner et al (1995) *Am. J. Med. Genet.* 57:333-337).

DNA Electrophoresis and Hybridization

DNA extraction, restriction digests, electrophoresis, Southern blotting, hybridization, and washing were performed by standard methods (Sambrook et al (1989) *Molecular Cloning: A Laboratory Manual. Second Edition Cold Spring Harbor Laboratory Press*).

cDNA Screening and Characterization

Exon amplification was carried out using the pSPL3 vector as described by Church et al 1994. A human fetal brain cDNA library in lambdaZAPII (Stratagene) was screened by standard methods using exon probes. cDNA clones and trapped exons were sequenced manually (Sanger et al (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467) with Sequenase T7 DNA polymerase (U.S. Biochemicals).

RNA Procedures

Cytoplasmic RNA was isolated by standard methods (Sambrook et al (1989) *Molecular Cloning: A Laboratory Manual, Second Edition Cold Spring Harbor Laboratory Press*) or using RNazol (Biogenesis, UK). RNA was reverse transcribed using oligo(dT) or

random hexamer primers and Superscript Reverse Transcriptase (Gibco). Portions of the cDNA were amplified using primer sets described in the text. Direct sequencing of PCR products was carried out as described (McClatchey et al (1992) *Cell* 68:769-774) or by purification with Qiagel (Qiagen) followed by sequencing with an ABI 373A automated sequencer. PCR products were subcloned using the TA Cloning Kit (Invitrogen).

Polymerase Chain Reaction

The polymerase chain reaction was carried out using Taq polymerase, following the recommendations of the manufacturer. The oligonucleotide primers used in the experiments are described in Table 1. The assay for the "56" deletion was carried out on 100 ng of genomic DNA using primers F2 (SEQ ID NO: 4) and P3 (SEQ ID NO: 11) (Table 1) in a reaction including 0.2 μM each primer, 0.2mM each dNTP, 1.5 mM $MgCl_2$ and 0.5-1 μl AmpliTaq (Perkins Elmer). In one laboratory, the reaction was supplemented with 5 units TaqExtender (Stratagene) which was found to enhance the amplification. Annealing temperatures ranging between 55°C and 62°C were used successfully. Samples were fractionated on an 0.8% agarose gel.

Genomic Sequencing

Genomic DNA from a normal control and the somatic cell hybrid CY101 which carries a single copy of chromosome 16 derived from a patient homozygous for the "56" haplotype was PCR amplified with primers P1-P3 (SEQ ID NOS: 3 and 11) (Table 1). The resulting PCR products were digested with TaqI. A 1.5 kb fragment was detected in the control and a 0.5 kb fragment was detected in CY101. These two fragments were subcloned into pUC19 and sequenced with an ABI 373A automated sequencer. In an independent study, the sequence spanning the "56" deletion was generated by PCR sequencing of the subcloned 3.8 kb PstI fragment using an ABI 373A automated sequencer.

Additional Mutations Disrupting the Batten Disease Gene

A PCR-based assay was used to screen for the 1.02 Kb deletion in the pooled Batten disease patient resource of 194 families. Fourteen individuals did not have the 1.02 Kb deletion whilst 41 were found to be heterozygous and 139 homozygous for this mutation. Thus, 55 individuals in our resource possessed other mutations, including three which have been described above.

To determine the range of mutations present in the 52 individuals carrying unknown mutations, we designed primers to amplify each exon of the gene and surrounding intron sequence and performed SSCP and direct sequencing analysis. A total of 15 sets of primers were used (Table 3). Nineteen novel mutations were found (Table 3, Figure 8): six missense, five nonsense, three small deletions, three small insertions, one intronic and one

splice site. An example of the delineation of a nonsense mutation is shown in Figure 10. In total, the mutations in 31/52 individuals were defined on both chromosomes and therefore, the disease-causing mutations in 89% (173/194) of the patients in the resource were delineated, making a total of 23 disease-causing mutations reported to date in *CLN3*.

A founder effect responsible for the 1.02 Kb deletion present in the majority of Batten patients, associated with the haplotype "56" for alleles at markers D16S299-D16S298 has been described herein. The majority of the newly described mutations are present in only one family, however, five occur in more than one family (Table 3). Examination of the families with the same mutation reveals each to have an identical or related haplotype suggesting the existence of smaller founder effects, with two (561delG/haplotype "44" and C1137T/haplotype "66") concentrated in the Dutch population, and three (1081insA/haplotype "63", G1138A/haplotype "45" and C1191T/haplotype "54") founded worldwide.

All six missense mutations in *CLN3* affect residues which are identical between the human and its homologues in *Saccharomyces cerevisiae* (*YHC3*) (accession number Z49334), dog (L76281) and mouse (U47106). Five out of the six residues are also conserved in the homologue in *Caenorhabditis elegans* (Z77656). A structural model for the Batten disease protein is proposed in Figure 9. Two residues affected by missense mutations are located in predicted transmembrane segments of the protein, four are located on predicted extracellular loops on one face only of the protein (three are in the same predicted loop) (Figure 9) suggesting that this face is particularly important for normal function. Two different missense mutations affect Arg334 indicating that this residue plays a critical role in the normal functioning of the CLN3 protein. The identification of such critical residues facilitates the determination of important structural and functional domains of the protein.

Out of the 52 patients who carried unknown mutations, mutations in 32 patients have been delineated with mutations on both chromosomes identified in 31. The twenty remaining patients where the mutation on one or both chromosomes is not known have been completely screened across all exons and surrounding intronic sequence suggesting that additional mutations lie either in the promoter region or elsewhere in an intron. Thirteen of these are heterozygous for the 1.02 Kb deletion and therefore almost certainly have Batten disease. However seven do not carry the 1.02 Kb deletion on a chromosome, so it is possible that they do not carry mutations in *CLN3*, although their clinical symptoms suggest Batten disease. Any mutations which remain undetected in this Batten patient resource may be found by applying other approaches such as Southern blotting, long range PCR and sequencing of the promoter region.

The novel mutations are outlined in Table 3 below.

Table 3 Novel mutations identified in *CLN3*

Family[a] Number	Haplotype[b]	Mutation	Nucleotide change	Amino acid change	Location[c]	Inheritance[d]	Restriction site change	Number of families with the mutation	Country of origin
L39	56*/46	Missense	T439C	Leu101Pro	Exon 5	Maternal	*Bsi*HKAI (loss)	1	The Netherlands
L1	44/44	1bp deletion	561delG	Frameshift after Leu141	Exon 6	Both	*Bst*NI (loss)	5	The Netherlands (3), USA (2)
L227	56*/54	2bp insertion	511ins CC	Frameshift after Pro126	Exon 6	Maternal	-	1	UK
L121/BA	56*/64	Intron change	598(-13G->C)	Aberrant splicing^ Truncated protein	Intron 6	Paternal	-	1	USA
L29	56*/66	Nonsense	C619G	Ser161STOP	Exon 7	Maternal	*Sau* 3A (gain)	1	Sweden
L259	45*/32	Nonsense	C622G	Ser162STOP	Exon 7	Maternal	*Nla*III (gain)	1	Denmark
L46	56*/64	Missense	T646C	Leu170Pro	Exon 7	Maternal	-	1	UK
L189	44[$]/34	Nonsense	C768T	Gln211STOP	Exon 8	Maternal	*Acc*I (gain)	1	Italy
L250	n3/n3	1bp insertion	723insG	Frameshift after Gly195	Exon 8	ND	-	1	UK
L116	66/66	2bp deletion	695delAG	Frameshift after Ser185	Exon 8	Both	*Alw*NI (loss)	1	Italy
L285	n6*/n6	Missense	G1020A	Glu295Lys	Exon 11	Maternal	-	1	Finland
L209	63/63	1bp insertion	1081insA	Frameshift after Ser314	Exon 12	Both	*Hinc*II (gain)	4	Italy (2), Iceland, USA
L10	56*/66	Missense	C1137T	Arg334Cys	Exon 13	Paternal	*Bsr*BI (loss)	3	The Netherlands (3)
L204	56*/45	Missense	G1138A	Arg334His	Exon 13	Paternal	*Bsr*BI (loss)	4	Finland, UK, Germany; USA
L216	56*/66	Missense	G1125T	Val330Phe	Exon 13	Maternal	-	1	Norway
L243	26*/43	Nonsense	C1116T	Gln327STOP	Exon 13	Maternal	*Bfa*I (gain)	1	Denmark

L8	56*/54	Nonsense	C1191T	Gln352STOP	Exon 13	Maternal	*PstI* (loss)	2	The Netherlands. USA
BB	56*/26	Splice site	1335(-1G->T)	Aberrant splicing^ Truncated protein	Intron 14	Maternal	-	1	USA
L61	56*/63	1bp deletion	1409delG	Frameshift after Ser423	Exon 15	ND	-	1	UK

[a]Details for the family in which the mutation was originally found are shown; [b]Haplotypes are formed by the markers *D16S299* and *D16S298*; [c]Exon numbering taken from Mitchison et al., *Genomics*, submitted; [d]Parents were checked for the novel mutation to confirm inheritance; *indicates a chromosome with the 1.02 Kb deletion; Bold lettering indicates the chromosome with the novel mutation; $indicates a chromosome for which the mutation is not yet identified; n indicates that the *D16S299* marker has not been typed; ND indicates that is was not possible to confirm the parental origin of the mutation. ^Aberrant splicing was confirmed using RT-PCR analysis and sequencing. None of the missense mutations are present on 90 normal chromosomes by sequencing. The PCR primers for amplification of *CLN3* exons are: **Exon 1** - (5'-aaaggtacaggcctcagggt-3')(SEQ ID NO:28) and (5'-agctctcattcccctcaggt-3')(SEQ ID NO:29); **Exon 2** - (5'-acctgagggaatgagagct-3')(SEQ ID NO:30) and (5'-tgggttcagctcctttgc-3')(SEQ ID NO:31); **Exon 3** - (5'-attgaagggcataggtaaga-3')(SEQ ID NO:32) and (5'-actttaccccaccttgtccc-3')(SEQ ID NO:33); **Exon 4** - (5'-tcaagtgaaggcagagctgg-3')(SEQ ID NO:34) and (5'-agtcccagctgggtagtgaa-3')(SEQ ID NO:35); **Exon 5** - (5'-cctgtgtttgtagcaggcct-3')(SEQ ID NO:36) and (5'-aaggtcggtctctactctcagc-3')(SEQ ID NO:37); **Exon 6** - (5'-tggtcaggagctgagaaagg-3')(SEQ ID NO:38) and (5'-gaatccctttcctctgggag-3')(SEQ ID NO:39); **Exon 7** - (5'-ggagcctctatgagctgatactg-3')(SEQ ID NO:40) and (5'-ggaacattcaggaggacctagg-3')(SEQ ID NO:41); **Exon 8** - (5'-tgtcccatggtcagcctag-3')(SEQ ID NO:42) and (5'-ttctctccttggacccctct-3'(SEQ ID NO:43); **Exon 9** - (5'-gcagtgagctacccatcttt-3')(SEQ ID NO:44) and (5'-aggaaaaggccaaacccag-3')(SEQ ID NO:45); **Exon 10** - (5'-aatccagtggcatggaagttg-3')(SEQ ID NO:46) and (5'-ctacgaccaagggaacaat-3')(SEQ ID NO:47) and (5'-ctacgaccaagggaacaat-3')(SEQ ID NO:48); **Exon 11** - (5'-tcgggaaaggtggacagt-3')(SEQ ID NO:49) and (5'-ggtattgctgagcgtgactc-3')(SEQ ID NO:50); **Exon 12** - (5'-tcgggaaaggtggacagt-3')(SEQ ID NO:49) and (5'-aggtgaaacggatgcgac-3')(SEQ ID NO:51); **Exon 13** - (5'-tttgaactcctcttttctgg-3')(SEQ ID NO:52) and (5'-acactttccactgatagtggga-3')(SEQ ID NO:53); **Exon 14** - (5'-tcctaaaaccagggacccct-3')(SEQ ID NO:54) and (5'-ttcagtcccagacatccctg-3')(SEQ ID NO:55); **Exon 15** - (5'- agggatgtctgggactgaag-3')(SEQ ID NO:56) and (5'- ggcatgatgccaggaaga-3')(SEQ ID NO:57).

Experimental Procedures

Families

One hundred and ninety four families with Batten disease from 20 countries were included in this study. A definition of classical Batten disease as onset of visual

disorder 6.2 ± 1.8 yrs, dementia at 7.4 ± 2 yrs, seizures and motor disturbance at 9.5 ± 3.5 yrs with onset of a vegetative state at 18.4 ± 2.8 yrs and mean age of death 20.2 ± 6.3 yrs was followed.

Genomic DNA was extracted directly from peripheral blood or from lymphoblastoid cell lines using standard methods.

1.02Kb deletion assay

Three PCR-based methods were used to detect the 1.02 Kb genomic deletion: Either primers F2(SEQ ID NO:4)/P3(SEQ ID NO:11) were used to amplify DNA surrounding the deletion or, where long-range PCR was not possible due to the age and quality of patient DNA, primers F2(SEQ ID NO:4)/R1(SEQ ID NO:10) or primers which amplify exon 7 (Table 3) were used to check the absence of exon 7. Positive controls for PCR of other *CLN3* exons were included. All results were concordant with the observed haplotypes for alleles at markers *D16S299* and *D16S298*.

PCR amplification of exons

Primers to amplify each exon and the surrounding intron sequence were designed from genomic DNA sequence of *CLN3*. PCR was performed in a final volume of 100 µl using 100 ng of genomic DNA, 0.2 µM of each primer, 0.25 mM of each dNTP, 1.5 mM $MgCl_2$ and 0.3 µl of AmpliTaq (Perkin-Elmer). A 'hot' start was performed followed by 1 min at 94°C, 1 min at 60°C, 1 min at 72°C (30 cycles), and 10 min at 72°C (1 cycle) using a Hybaid OmniGene. The resulting products were electrophoresed in 1% agarose gels and were visualized after ethidium bromide staining with a UV transilluminator.

SSCP

Two different systems were used for the detection of single strand conformational polymorphisms (SSCP). The first used the Phastsystem (Pharmacia). Gels were electrophoresed for 300Vhr at 4°C and for 200Vhr at 15°C in this study. The second method used a radioactive protocol and samples were analyzed on MDE™ high-resolution gels (AT Biochem).

Direct DNA sequencing

Amplified exon products to be sequenced were desalted/concentrated using a Microcon-100 column (Amicon). Sequencing was carried out with the same primers used for exon amplification using the Taq FS Dye Terminator Cycle sequencing kit (Perkin-Elmer) and automated analysis was done with the ABI 373A sequencer. Sequence comparisons were performed using Sequence Navigator software (Perkin-Elmer). The exons were sequenced manually with Sequenase T7 DNA polymerase (United States Biochemicals).

RNA extraction and analysis

Cytoplasmic RNA was isolated using standard methods. RNA was reverse transcribed using oligo(dT) and Superscript reverse transcriptase (Gibco-BRL). Primers 6795 (5'-ttgatccttgtcacctgtcg-3') and 6797 (5'-attcagaaggcatgatgcc-3') were used to amplify the RNA-cDNA duplex from patient BA, followed by ampification using nested primers 6972 (5'-aaattgttggctcctcttgg-3') and 6333 (5'-ggctgggagcacagttcat-3'). Primers 6972 and 6700 (5'-gcgctctctgcttcttcttc-3') were used to amplify the RNA-cDNA duplex from patient L121/BB. All products were subcloned and sequenced.

Restriction endonuclease analysis

Amplified exon products were digested according to the manufacturer's recommendations. Samples were electrophoresed in 1% agarose gels and were visualized after ethidium bromide staining with a UV transilluminator.

Isolation of CLN3 homologs

One of ordinary skill in the art can apply routine methods to obtain CLN3 homologs, e.g., CLN3 genes from different species. For example, degenerate oligonucleotide primers can be synthesized from the regions of homology shared by human and mouse CLN3 genes. The degree of degeneracy of the primers will depend on the degeneracy of the genetic code for that particular amino acid sequence used. The degenerate primers should also contain restriction endonuclease sites at the 5' end to facilitate subsequent cloning.

Total mRNA can be obtained from cells, e.g., brain cells, and reverse transcribed using Superscript Reverse Transcriptase Kit. Instead of an oligo(dT) primer supplied with the kit, one can use one of the 3' degenerate oligonucleotide primers to increase the specificity of the reaction. After a first strand synthesis, cDNA obtained can than be subjected to a PCR amplification using above described degenerate oligonucleotides. PCR conditions should be optimized for the annealing temperature, Mg^{++} concentration and cycle duration.

Once the fragment of appropriate size is amplified, it should be Klenow filled, cut with appropriate restriction enzymes and gel purified. Such fragment can than be cloned into a vector, e.g., a Bluescript vector. Clones with inserts of appropriate size can be digested with restriction enzymes to compare generated fragments with those of other CLN3 genes, e.g., hauman and mouse CLN3 genes. Those clones with distinct digestion profiles can be sequenced.

Alternatively, antibodies can be made to the conserved regions of the human and/or mouse CLN3 genes and used to screen expression libraries.

Gene Therapy

The gene constructs of the invention can also be used as a part of a gene therapy protocol to deliver nucleic acids encoding either an agonistic or antagonistic form of a Batten disease polypeptide. The invention features expression vectors for *in vivo* transfection and expression of a Batten disease polypeptide in particular cell types (e.g., neural cells) so as to reconstitute the function of, enhance the function of, or alternatively, antagonize the function of a Batten disease polypeptide in a cell in which the polypeptide is misexpressed.

Expression constructs of Batten disease polypeptides, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the Batten disease gene to cells *in vivo*. Approaches include insertion of the subject gene into viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA encoding a Batten disease polypeptide. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors can be used as a recombinant gene delivery system for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). A replication defective retrovirus can be packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψCrip, ψ

Cre, ψ2 and ψAm. Retroviruses have been used to introduce a variety of genes into many different cell types, including epithelial cells, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464; Wilson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3014-3018; Armentano et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6141-6145; Huber et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8039-8043; Ferry et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8377-8381; Chowdhury et al. (1991) *Science* 254:1802-1805; van Beusechem et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:7640-7644; Kay et al. (1992) *Human Gene Therapy* 3:641-647; Dai et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10892-10895; Hwu et al. (1993) *J. Immunol.* 150:4104-4115; U.S. Patent No. 4,868,116; U.S. Patent No. 4,980,286; PCT Application WO 89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

Another viral gene delivery system useful in the present invention utilizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See, for example, Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain circumstances in that they are not capable of infecting nondividing cells and can be used to infect a wide variety of cell types, including epithelial cells (Rosenfeld et al. (1992) cited *supra*). Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis in situations where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al. cited *supra*; Haj-Ahmand and Graham (1986) *J. Virol.* 57:267).

Yet another viral vector system useful for delivery of the subject Batten disease gene is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol.*

Biol. 7:349-356; Samulski et al. (1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J. Virol.* 62:1963-1973). Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector such as that described in Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260 can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see for example Hermonat et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470; Tratschin et al. (1985) *Mol. Cell. Biol.* 4:2072-2081; Wondisford et al. (1988) *Mol. Endocrinol.* 2:32-39; Tratschin et al. (1984) *J. Virol.* 51:611-619; and Flotte et al. (1993) *J. Biol. Chem.* 268:3781-3790).

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a Batten disease polypeptide in the tissue of a mammal, such as a human. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject Batten disease gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In a representative embodiment, a gene encoding a Batten disease polypeptide can be entrapped in liposomes bearing positive charges on their surface (e.g., lipofectins) and (optionally) which are tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al. (1992) *No Shinkei Geka* 20:547-551; PCT publication WO91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075).

In clinical settings, the gene delivery systems for the therapeutic Batten disease gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) *PNAS* 91: 3054-3057). In a preferred embodiment of the invention, the Batten disease gene is targeted to neural cells.

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced in tact from recombinant cells, e.g. retroviral

vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

Antisense Therapy

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotides or their derivatives which specifically hybridize (e.g. bind) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a Batten disease polypeptide, or mutant thereof, so as to inhibit expression of the encoded protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

In one embodiment, the antisense construct binds to a naturally-occurring sequence of a Batten disease gene which, for example, is involved in expression of the gene. These sequences include, for example, start codons, stop codons, and RNA primer binding sites.

In another embodiment, the antisense construct binds to a nucleotide sequence which is not present in the wild type gene. For example, the antisense construct can bind to a region of a Batten disease gene which contains an insertion of an exogenous, non-wild type sequence. Alternatively, the antisense construct can bind to a region of a Batten disease gene which has undergone a deletion, thereby bringing two regions of the gene together which are not normally positioned together and which, together, create a non-wild type sequence.

When administered *in vivo* to a subject, antisense constructs which bind to non-wild type sequences provide the advantage of inhibiting the expression of mutant Batten disease genes (e.g., which encode polypeptides which are unstable, have an undesirable activity, or otherwise give rise to disorders associated with Batten disease), without inhibiting expression of any wild type Batten disease gene

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a Batten disease polypeptide. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a Batten disease gene. Such oligonucleotide probes are preferably modified oligonucleotide which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are

phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included in the invention.

The compounds can be administered orally, or by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives, and detergents. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind.

The antisense constructs of the present invention, by antagonizing the expression of a Batten disease gene, can be used in the manipulation of tissue, both *in vivo* and in *ex vivo* tissue cultures.

Transgenic Animals

The invention includes transgenic animals which include cells (of that animal) which contain a Batten disease transgene and which preferably (though optionally) express (or misexpress) an endogenous or exogenous Batten disease gene in one or more cells in the animal.

The Batten disease transgene can encode a mutant Batten disease polypeptide, thereby creating an animal model for Batten disease. Such animals can be used as disease

models or can be used to screen for agents effective at treating Batten disease. Alternatively, the Batten disease transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, or tissues utilizing, for example, cis-acting sequences that control expression in the desired pattern. Tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences. In preferred embodiments, the transgenic animal carries a "knockout" Batten disease gene, i.e., a deletion of all or a part of the gene.

Genetic techniques which allow for the expression of transgenes, that are regulated *in vivo* via site-specific genetic manipulation, are known to those skilled in the art. For example, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of the subject Batten disease gene. For example, excision of a target sequence which interferes with the expression of a recombinant Batten disease gene, such as one which encodes an agonistic homolog, can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the Batten disease gene from the promoter element or an internal stop codon.

Moreover, the transgene can be made so that the coding sequence of the gene is flanked with recombinase recognition sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation. See e.g., descriptions of the *cre/loxP* recombinase system of bacteriophage P1 (Lakso et al. (1992) *PNAS* 89:6232-6236; Orban et al. (1992) *PNAS* 89:6861-6865) or the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355; PCT publication WO 92/15694). Genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element.

Thus, the activation expression of the recombinant Batten disease gene can be regulated via control of recombinase expression.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneous expressed in order to facilitate expression of the transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080. Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, the Batten disease transgene could remain silent into adulthood until "turned on" by the introduction of the trans-activator.

Production of Fragments and Analogs

The inventor has provided the primary amino acid structure of a Batten disease polypeptide. Once an example of this core structure has been provided, one skilled in the art can alter the disclosed structure by producing fragments or analogs, and testing the newly produced structures for activity. Examples of prior art methods which allow the production and testing of fragments and analogs are discussed below. These, or analogous methods can be used to make and screen fragments and analogs of a Batten disease polypeptide having at least one biological activity e.g., which react with an antibody (e.g., a monoclonal antibody) specific for a Batten disease polypeptide.

Generation of Fragments

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNA's which encode an array of fragments. DNA's which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

Production of Altered DNA and Peptide Sequences: Random Methods

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein.)

PCR Mutagenesis

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). This is a very powerful and relatively rapid method of introducing random mutations. The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn^{2+} to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

Saturation Mutagenesis

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complementary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

Degenerate Oligonucleotides

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed

in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Production of Altered DNA and Peptide Sequences: Methods for Directed Mutagenesis

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions. of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site. or combinations of options 1-3.

Alanine Scanning Mutagenesis

Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

Oligonucleotide-Mediated Mutagenesis

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA* 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or

bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci.* USA, 75: 5765[1978]). For purposes of the present invention, preferred oligonucleotide primers have a nucleotide sequence shown in SEQ ID NOS: 3-15.

Cassette Mutagenesis

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the protein subunit DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

Combinatorial Mutagenesis

Combinatorial mutagenesis can also be used to generate mutants, e.g., a library of variants which is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

Primary High-Through-Put Methods for Screening Libraries of Peptide Fragments or Homologs

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to an antibody specific for a Batten disease polypeptide. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

Display Libraries

In one approach to screening assays, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homolog which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over 10^{13} phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH_2-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines 91*, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of may peptides copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane protease IgA of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull *et al.* (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all

related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89-1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7-10^9 independent clones are routinely prepared. Libraries as large as 10^{11} recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10^{12} decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be

affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret. et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

Secondary Screens

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of a protein of interest is identified, such as the primary amino acid sequence of a Batten disease polypeptide as disclosed herein, it is routine to perform for one skilled in the art to obtain analogs and fragments.

Antibodies

The invention also includes antibodies specifically reactive with a subject Batten disease polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject Batten disease polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for

antigenic determinants of the Batten disease polypeptide of the invention, e.g. antigenic determinants of a polypeptide of
SEQ ID NO: 2.

The term "antibody", as used herein, intended to include fragments thereof which are also specifically reactive with a Batten disease polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having an anti-Batten disease polypeptide portion.

Both monoclonal and polyclonal antibodies (Ab) directed against Batten disease polypeptides, or fragments or analogs thereof, and antibody fragments such as Fab` and $F(ab`)_2$, can be used to block the action of a Batten disease polypeptide and allow the study of the role of a Batten disease polypeptide of the present invention.

Antibodies which specifically bind Batten disease polypeptide epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of Batten disease polypeptide. Anti-Batten disease polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate wild type or mutant Batten disease polypeptide levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor Batten disease polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with disorders associated with Batten disease. The level of a Batten disease polypeptide can be measured in cells found in bodily fluid, such as in samples of cerebral spinal fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-Batten disease polypeptide antibodies can include, for example, immunoassays designed to aid in early diagnosis of Batten disease polypeptide-mediated disorders, e.g., to detect cells in which a mutation of the Batten disease gene has occurred.

Another application of anti-Batten disease antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject Batten disease polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-Batten disease polypeptide antibodies. Phage, scored by this assay, can then be isolated

from the infected plate. Thus, the presence of Batten disease homologs can be detected and cloned from other animals, and alternate isoforms (including splicing variants) can be detected and cloned from human sources.

Drug Screening Assays

By making available purified and recombinant-Batten disease polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject Batten disease polypeptide. In one embodiment, the assay evaluates the ability of a compound to modulate binding between a Batten disease polypeptide and a naturally occurring ligand, e.g., an antibody specific for a Batten disease polypeptide. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target.

Other Embodiments

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide of SEQ ID NO:2 (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and, polypeptides specifically bound by antisera to a Batten disease polypeptide.

The invention also includes fragments, preferably biologically active fragments, or analogs of a Batten disease polypeptide. A biologically active fragment or analog is one having any *in vivo* or *in vitro* activity which is characteristic of the Batten disease polypeptide shown in SEQ ID NO:2, or of other naturally occurring Batten disease polypeptides, e.g., one or more of the biological activities described above. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's.

Fragments include those expressed in native or endogenous cells, e.g., as a result of post-translational processing, e.g., as the result of the removal of an amino-terminal signal sequence, as well as those made in expression systems, e.g., in CHO cells. Because peptides, such as a Batten disease polypeptide, often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful Batten disease polypeptide fragment or Batten disease polypeptide analog is one which exhibits a biological activity in any biological assay for Batten disease polypeptide activity. Most preferably the fragment or analog possesses 10%, preferably 40%, or at least 90% of the activity of a Batten disease polypeptide (SEQ ID NO: 2), in any *in vivo* or *in vitro* Batten disease polypeptide activity assay.

Analogs can differ from a naturally occurring Batten disease polypeptide in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include *in vivo* or *in vitro* chemical derivatization of a Batten disease polypeptide. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation.

Preferred analogs include a Batten disease polypeptide (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not abolish the Batten disease polypeptide biological activity. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be taken from the table below.

TABLE 4

CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln

Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β-Ala Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu. D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn. D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val. D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile. D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs

that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to a Batten disease polypeptide analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues, preferably at least about 60 residues in length. Fragments of a Batten disease polypeptide can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of a Batten disease polypeptide can be assessed by methods known to those skilled in the art, as described herein. Also included are Batten disease polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing events.

In order to obtain a Batten disease polypeptide, a Batten disease polypeptide-encoding DNA can be introduced into an expression vector, the vector introduced into a cell suitable for expression of the desired protein, and the peptide recovered and purified, by prior art methods. Antibodies to the peptides an proteins can be made by immunizing an animal, e.g., a rabbit or mouse, and recovering anti-Batten disease polypeptide antibodies by prior art methods.

Equivalents

Those skilled in the art will be able to recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Massachusetts General Hospital
Molecular Neurogenetics Unit
(B) STREET: Thirteenth Street
(C) CITY: Charlestown
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02129

(A) NAME: Leiden University Institutional Development
Department of Human Genetics
(B) STREET: Wassenaarseweg 72
(C) CITY: Leiden
(D) STATE:
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): 2333 A1

(A) NAME: University College London Medical School
Department of Pediatrics, The Rayne Institute
(B) STREET: University Street
(C) CITY: London
(D) STATE:
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): WC1E 6JJ

(ii) TITLE OF INVENTION: Batten Disease Gene

(iii) NUMBER OF SEQUENCES: 58

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(vii) PRIOR APPLICATION DATA:
(A) PROVISIONAL APPLICATION SERIAL NUMBER: 60/003,030
(B) FILING DATE: 31-AUG-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Myers, Louis
(B) REGISTRATION NUMBER: 35,965
(C) REFERENCE/DOCKET NUMBER: MGP-035PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)227-5941

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 138..1451

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CCCCTAGACA AGCCGGAGCT GGGACCGGCA ATCGGGCGTT GATCCTTGTC ACCTGTCGCA      60

GACCCTCATC CCTCCCGTGG GAGCCCCCTT TGGACACTCT ATGACCCTGG ACCCTCGGGG     120

GACCTGAACT TGATGCG ATG GGA GGC TGT GCA GGC TCG CGG CGG CGC TTT         170
                   Met Gly Gly Cys Ala Gly Ser Arg Arg Arg Phe
                     1               5                  10

TCG GAT TCC GAG GGG GAG GAG ACC GTC CCG GAG CCC CGG CTC CCT CTG       218
Ser Asp Ser Glu Gly Glu Glu Thr Val Pro Glu Pro Arg Leu Pro Leu
               15                  20                  25

TTG GAC CAT CAG GGC GCG CAT TGG AAG AAC GCG GTG GGC TTC TGG CTG       266
Leu Asp His Gln Gly Ala His Trp Lys Asn Ala Val Gly Phe Trp Leu
          30                  35                  40

CTG GGC CTT TGC AAC AAC TTC TCT TAT GTG GTG ATG CTG AGT GCC GCC       314
Leu Gly Leu Cys Asn Asn Phe Ser Tyr Val Val Met Leu Ser Ala Ala
      45                  50                  55

CAC GAC ATC CTT AGC CAC AAG AGG ACA TCG GGA AAC CAG AGC CAT GTG       362
His Asp Ile Leu Ser His Lys Arg Thr Ser Gly Asn Gln Ser His Val
  60                  65                  70                  75

GAC CCA GGC CCA ACG CCG ATC CCC CAC AAC AGC TCA TCA CGA TTT GAC       410
Asp Pro Gly Pro Thr Pro Ile Pro His Asn Ser Ser Ser Arg Phe Asp
                  80                  85                  90

TGC AAC TCT GTC TCT ACG GCT GCT GTG CTC CTG GCG GAC ATC CTC CCC       458
Cys Asn Ser Val Ser Thr Ala Ala Val Leu Leu Ala Asp Ile Leu Pro
              95                  100                 105

ACA CTC GTC ATC AAA TTG TTG GCT CCT CTT GGC CTT CAC CTG CTG CCC       506
Thr Leu Val Ile Lys Leu Leu Ala Pro Leu Gly Leu His Leu Leu Pro
         110                 115                 120

TAC AGC CCC CGG GTT CTC GTC AGT GGG ATT TGT GCT GCT GGA AGC TTC       554
Tyr Ser Pro Arg Val Leu Val Ser Gly Ile Cys Ala Ala Gly Ser Phe
     125                 130                 135
```

```
GTC CTG GTT GCC TTT TCT CAT TCT GTG GGG ACC AGC CTG TGT GGT GTG      602
Val Leu Val Ala Phe Ser His Ser Val Gly Thr Ser Leu Cys Gly Val
140                 145                 150                 155

GTC TTC GCT AGC ATC TCA TCA GGC CTT GGG GAG GTC ACC TTC CTC TCC      650
Val Phe Ala Ser Ile Ser Ser Gly Leu Gly Glu Val Thr Phe Leu Ser
                160                 165                 170

CTC ACT GCC TTC TAC CCC AGG GCC GTG ATC TCC TGG TGG TCC TCA GGG      698
Leu Thr Ala Phe Tyr Pro Arg Ala Val Ile Ser Trp Trp Ser Ser Gly
            175                 180                 185

ACT GGG GGA GCT GGG CTG CTG GGG GCC CTG TCC TAC CTG GGC CTC ACC      746
Thr Gly Gly Ala Gly Leu Leu Gly Ala Leu Ser Tyr Leu Gly Leu Thr
        190                 195                 200

CAG GCC GGC CTC TCC CCT CAG CAG ACC CTG CTG TCC ATG CTG GGT ATC      794
Gln Ala Gly Leu Ser Pro Gln Gln Thr Leu Leu Ser Met Leu Gly Ile
    205                 210                 215

CCT GCC CTG CTG CTG GCC AGC TAT TTC TTG TTG CTC ACA TCT CCT GAG      842
Pro Ala Leu Leu Leu Ala Ser Tyr Phe Leu Leu Leu Thr Ser Pro Glu
220                 225                 230                 235

GCC CAG GAC CCT GGA GGG GAA GAA GAA GCA GAG AGC GCA GCC CGG CAG      890
Ala Gln Asp Pro Gly Gly Glu Glu Glu Ala Glu Ser Ala Ala Arg Gln
                240                 245                 250

CCC CTC ATA AGA ACC GAG GCC CCG GAG TCG AAG CCA GGC TCC AGC TCC      938
Pro Leu Ile Arg Thr Glu Ala Pro Glu Ser Lys Pro Gly Ser Ser Ser
            255                 260                 265

AGC CTC TCC CTT CGG GAA AGG TGG ACA GTA TTC AAG GGT CTG CTG TGG      986
Ser Leu Ser Leu Arg Glu Arg Trp Thr Val Phe Lys Gly Leu Leu Trp
        270                 275                 280

TAC ATT GTT CCC TTG GTC GTA GTT TAC TTT GCC GAG TAT TTC ATT AAC     1034
Tyr Ile Val Pro Leu Val Val Val Tyr Phe Ala Glu Tyr Phe Ile Asn
    285                 290                 295

CAG GGA CTT TTT GAA CTC CTC TTT TTC TGG AAC ACT TCC CTG AGT CAC     1082
Gln Gly Leu Phe Glu Leu Leu Phe Phe Trp Asn Thr Ser Leu Ser His
300                 305                 310                 315

GCT CAG CAA TAC CGC TGG TAC CAG ATG CTG TAC CAG GCT GGC GTC TTT     1130
Ala Gln Gln Tyr Arg Trp Tyr Gln Met Leu Tyr Gln Ala Gly Val Phe
                320                 325                 330

GCC TCC CGC TCT TCT CTC CGC TGC TGT CGC ATC CGT TTC ACC TGG GCC     1178
Ala Ser Arg Ser Ser Leu Arg Cys Cys Arg Ile Arg Phe Thr Trp Ala
            335                 340                 345

CTG GCC CTG CTG CAG TGC CTC AAC CTG GTG TTC CTG CTG GCA GAC GTG     1226
Leu Ala Leu Leu Gln Cys Leu Asn Leu Val Phe Leu Leu Ala Asp Val
        350                 355                 360

TGG TTC GGC TTT CTG CCA AGC ATC TAC CTC GTC TTC CTG ATC ATT CTG     1274
```

```
Trp Phe Gly Phe Leu Pro Ser Ile Tyr Leu Val Phe Leu Ile Ile Leu
    365                 370                 375

TAT GAG GGG CTC CTG GGA GGC GCA GCC TAC GTG AAC ACC TTC CAC AAC      1322
Tyr Glu Gly Leu Leu Gly Gly Ala Ala Tyr Val Asn Thr Phe His Asn
380                 385                 390                 395

ATC GCC CTG GAG ACC AGT GAT GAG CAC CGG GAG TTT GCA ATG GCG GCC      1370
Ile Ala Leu Glu Thr Ser Asp Glu His Arg Glu Phe Ala Met Ala Ala
                400                 405                 410

ACC TGC ATC TCT GAC ACA CTG GGG ATC TCC CTG TCG GGG CTC CTG GCT      1418
Thr Cys Ile Ser Asp Thr Leu Gly Ile Ser Leu Ser Gly Leu Leu Ala
            415                 420                 425

TTG CCT CTG CAT GAC TTC CTC TGC CAG CTC TCC TGATACTCGG GATCCTCAGG    1471
Leu Pro Leu His Asp Phe Leu Cys Gln Leu Ser
        430                 435

ACGCAGGTCA CATTCACCTG TGGGCAGAGG GACAGTCAGA CACCCAGGCC CACCCCAGAG    1531

ACCCTCCATG AACTGTGCTC CCAGCCTTCC CGGCAGGTCT GGGAGTAGGG AAGGGCTGAA    1591

GCCTTGTTTC CTTGCAGGGG GGCCAGCCAT TGTCTCCCAC TTGGGGAGTT TCTTCCTGGC    1651

ATCATGCCTT CTGAATAAAT GCCGATTTTG TCCATGGAAA AAAAAAAAAA AAAAAAAAAA    1711

AAAAAAAAAA AAAAAAAAAA A                                               1732
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 438 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Gly Gly Cys Ala Gly Ser Arg Arg Arg Phe Ser Asp Ser Glu Gly
 1               5                  10                  15

Glu Glu Thr Val Pro Glu Pro Arg Leu Pro Leu Leu Asp His Gln Gly
            20                  25                  30

Ala His Trp Lys Asn Ala Val Gly Phe Trp Leu Leu Gly Leu Cys Asn
        35                  40                  45

Asn Phe Ser Tyr Val Val Met Leu Ser Ala Ala His Asp Ile Leu Ser
    50                  55                  60

His Lys Arg Thr Ser Gly Asn Gln Ser His Val Asp Pro Gly Pro Thr
65                  70                  75                  80

Pro Ile Pro His Asn Ser Ser Ser Arg Phe Asp Cys Asn Ser Val Ser
                85                  90                  95
```

```
Thr Ala Ala Val Leu Leu Ala Asp Ile Leu Pro Thr Leu Val Ile Lys
            100                 105                 110

Leu Leu Ala Pro Leu Gly Leu His Leu Leu Pro Tyr Ser Pro Arg Val
        115                 120                 125

Leu Val Ser Gly Ile Cys Ala Ala Gly Ser Phe Val Leu Val Ala Phe
    130                 135                 140

Ser His Ser Val Gly Thr Ser Leu Cys Gly Val Val Phe Ala Ser Ile
145                 150                 155                 160

Ser Ser Gly Leu Gly Glu Val Thr Phe Leu Ser Leu Thr Ala Phe Tyr
                165                 170                 175

Pro Arg Ala Val Ile Ser Trp Trp Ser Ser Gly Thr Gly Gly Ala Gly
            180                 185                 190

Leu Leu Gly Ala Leu Ser Tyr Leu Gly Leu Thr Gln Ala Gly Leu Ser
        195                 200                 205

Pro Gln Gln Thr Leu Leu Ser Met Leu Gly Ile Pro Ala Leu Leu Leu
    210                 215                 220

Ala Ser Tyr Phe Leu Leu Leu Thr Ser Pro Glu Ala Gln Asp Pro Gly
225                 230                 235                 240

Gly Glu Glu Glu Ala Glu Ser Ala Ala Arg Gln Pro Leu Ile Arg Thr
                245                 250                 255

Glu Ala Pro Glu Ser Lys Pro Gly Ser Ser Ser Ser Leu Ser Leu Arg
            260                 265                 270

Glu Arg Trp Thr Val Phe Lys Gly Leu Leu Trp Tyr Ile Val Pro Leu
        275                 280                 285

Val Val Val Tyr Phe Ala Glu Tyr Phe Ile Asn Gln Gly Leu Phe Glu
    290                 295                 300

Leu Leu Phe Phe Trp Asn Thr Ser Leu Ser His Ala Gln Gln Tyr Arg
305                 310                 315                 320

Trp Tyr Gln Met Leu Tyr Gln Ala Gly Val Phe Ala Ser Arg Ser Ser
                325                 330                 335

Leu Arg Cys Cys Arg Ile Arg Phe Thr Trp Ala Leu Ala Leu Leu Gln
            340                 345                 350

Cys Leu Asn Leu Val Phe Leu Leu Ala Asp Val Trp Phe Gly Phe Leu
        355                 360                 365

Pro Ser Ile Tyr Leu Val Phe Leu Ile Ile Leu Tyr Glu Gly Leu Leu
    370                 375                 380

Gly Gly Ala Ala Tyr Val Asn Thr Phe His Asn Ile Ala Leu Glu Thr
385                 390                 395                 400
```

Ser Asp Glu His Arg Glu Phe Ala Met Ala Ala Thr Cys Ile Ser Asp
405 410 415

Thr Leu Gly Ile Ser Leu Ser Gly Leu Leu Ala Leu Pro Leu His Asp
420 425 430

Phe Leu Cys Gln Leu Ser
435

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTGATCCTTG TCACCTGTCG 20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TTCGTCCTGG TTGCCTTT 18

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGATCTCCTG GTGGTCCTCA 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGTCCATGCT GGGTATCCCT 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAAGAAGAAG CAGAGAGCGC 20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CAGCCCCTCA TAAGAACCGA 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGACGCAGGT CACATTCA 18

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGTGAGGGAG AGGAAGGTGA 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGCTCTCTGC TTCTTCTTCC 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTTGGCAGAA AGCCGAAC 18

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCCCTGCAAG GAAACAAG 18

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGCATGATGC CAGGAAAGA 19

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATTCAGAAGG CATGATGCC 19

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..217

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
 GT GTG GTC TTC GCT AGC ATC TCA TCA GGC CTT GGG GAG GTC ACC TTC        47
Gly Val Val Phe Ala Ser Ile Ser Ser Gly Leu Gly Glu Val Thr Phe
  1               5                  10                  15

CTC TCC CTC ACT GCC TTC TAC CCC AGG GCC GTG ATC TCC TGG TGG TCC        95
Leu Ser Leu Thr Ala Phe Tyr Pro Arg Ala Val Ile Ser Trp Trp Ser
            20                  25                  30

TCA GGG ACT GGG GGA GCT GGG CTG CTG GGG GCC CTG TCC TAC CTG GGC       143
Ser Gly Thr Gly Gly Ala Gly Leu Leu Gly Ala Leu Ser Tyr Leu Gly
        35                  40                  45

CTC ACC CAG GCC GGC CTC TCC CCT CAG CAG ACC CTG CTG TCC ATG CTG       191
Leu Thr Gln Ala Gly Leu Ser Pro Gln Gln Thr Leu Leu Ser Met Leu
    50                  55                  60

GGT ATC CCT GCC CTG CTG CTG GCC AG                                    217
Gly Ile Pro Ala Leu Leu Leu Ala Ser
65                  70
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
CCTGTGTGCT ATTTC                                                      15
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1658 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 142..1454

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
AATTCCGACA GCGGAACCTG GGACTGACCG CGGGGCATTG ATCCTTCGCA CCCACCTGTC      60

CCAGACTTTA ATCTGTTTTC TTGAAGCTAG CTCGGAACAC ACGCTGACTT TGGGCCCTTT     120

GGGGGACCCG AACTCAATGT T ATG GGA AGT TCT GCG GGC TCG TGG AGG CGC        171
                        Met Gly Ser Ser Ala Gly Ser Trp Arg Arg
                          1               5                  10

CTT GAG GAT TCT GAG AGG GAG GAG ACC GAC TCA GAG CCC CAG GCC CCT        219
Leu Glu Asp Ser Glu Arg Glu Glu Thr Asp Ser Glu Pro Gln Ala Pro
                    15                  20                  25

CGG TTG GAT AGT CGG AGT GTC CTT TGG AAG AAT GCA GTG GGT TTC TGG        267
Arg Leu Asp Ser Arg Ser Val Leu Trp Lys Asn Ala Val Gly Phe Trp
               30                  35                  40

ATC TTG GGT CTT TGC AAC AAT TTC TCA TAT GTG GTG ATG CTG AGC GCT        315
Ile Leu Gly Leu Cys Asn Asn Phe Ser Tyr Val Val Met Leu Ser Ala
          45                  50                  55

GCC CAT GAC ATC CTC AAG CAG GAG CAG GCG TCT GGA AAC CAG AGC CAT        363
Ala His Asp Ile Leu Lys Gln Glu Gln Ala Ser Gly Asn Gln Ser His
     60                  65                  70

GTA GAA CCA GGC CGA ACA CCC ACA CCC CAC AAC AGC TCA TCT CGA TTT        411
Val Glu Pro Gly Arg Thr Pro Thr Pro His Asn Ser Ser Ser Arg Phe
 75                  80                  85                  90

GAC TGC AAC TCC ATC TCC ACA GCT GCG GTG CTC CTA GCA GAC ATC CTT        459
Asp Cys Asn Ser Ile Ser Thr Ala Ala Val Leu Leu Ala Asp Ile Leu
                95                 100                 105

CCC ACC CTT GTC ATC AAA CTC CTG GCG CCT CTT GGC CTT CAC TTG CTG        507
Pro Thr Leu Val Ile Lys Leu Leu Ala Pro Leu Gly Leu His Leu Leu
           110                 115                 120

CCT TAC AGC CCC CGG GTG CTC GTC AGT GGA GTT TGT TCT GCT GGG AGC        555
Pro Tyr Ser Pro Arg Val Leu Val Ser Gly Val Cys Ser Ala Gly Ser
      125                 130                 135

TTT GTT CTG GTT GCC TTC TCT CAG TCA GTG GGG TTA AGC CTG TGT GGA        603
Phe Val Leu Val Ala Phe Ser Gln Ser Val Gly Leu Ser Leu Cys Gly
 140                 145                 150

GTG GTT TTG GCC AGC ATC TCC TCA GGG CTA GGG GAG GTC ACC TTC CTC        651
Val Val Leu Ala Ser Ile Ser Ser Gly Leu Gly Glu Val Thr Phe Leu
155                 160                 165                 170

TCA CTG ACT GCC TTC TAC CCC AGT GCT GTG ATC TCA TGG TGG TCT TCG        699
Ser Leu Thr Ala Phe Tyr Pro Ser Ala Val Ile Ser Trp Trp Ser Ser
                175                 180                 185

GGT ACC GGG GGT GCA GGG CTT CTT GGA TCG CTG TCT TAC CTG GGA CTC        747
Gly Thr Gly Gly Ala Gly Leu Leu Gly Ser Leu Ser Tyr Leu Gly Leu
           190                 195                 200

ACC CAG GCT GGC CTC TCC CCG CAG CAC ACC CTA CTT TCT ATG TTG GGG        795
Thr Gln Ala Gly Leu Ser Pro Gln His Thr Leu Leu Ser Met Leu Gly
```

```
          205                 210                 215

ATC CCT GTT CTG CTG CTA GCC AGC TAT TTC TTG TTG CTC ACG TCT CCT      843
Ile Pro Val Leu Leu Leu Ala Ser Tyr Phe Leu Leu Leu Thr Ser Pro
    220                 225                 230

GAA CCC TGG GAC CCT GGA GGA GAA AAC GAG GCA GAG ACT GCT GCC CGG      891
Glu Pro Trp Asp Pro Gly Gly Glu Asn Glu Ala Glu Thr Ala Ala Arg
235                 240                 245                 250

CAG CCT CTC ATA GGC ACC GAG ACC CCA GAG TCA AAG CCA GGT GCC AGC      939
Gln Pro Leu Ile Gly Thr Glu Thr Pro Glu Ser Lys Pro Gly Ala Ser
                255                 260                 265

TGG GAC CTC TCC CTC CAG GAA AGG TGG ACA GTG TTC AAG GGT CTC TTG      987
Trp Asp Leu Ser Leu Gln Glu Arg Trp Thr Val Phe Lys Gly Leu Leu
            270                 275                 280

TGG TAC ATC ATC CCT CTG GTG CTG GTC TAC TTT GCA GAA TAC TTT ATC     1035
Trp Tyr Ile Ile Pro Leu Val Leu Val Tyr Phe Ala Glu Tyr Phe Ile
        285                 290                 295

AAC CAG GGA CTT TTC GAG CTC CTG TTT TTC CGG AAC ACT TCC CTA AGC     1083
Asn Gln Gly Leu Phe Glu Leu Leu Phe Phe Arg Asn Thr Ser Leu Ser
    300                 305                 310

CAT GCT CAC GAG TAC CGA TGG TAC CAG ATG CTA TAC CAG GCT GGT GTG     1131
His Ala His Glu Tyr Arg Trp Tyr Gln Met Leu Tyr Gln Ala Gly Val
315                 320                 325                 330

TTC GCC TCC CGC TCT TCT CTC CAA TGT TGC CGA ATA CGG TTC ACC TGG     1179
Phe Ala Ser Arg Ser Ser Leu Gln Cys Cys Arg Ile Arg Phe Thr Trp
                335                 340                 345

GTC CTA GCC CTG CTC CAG AGC CTC AAC CTG GCC CTC CTG CTG GCA GAT     1227
Val Leu Ala Leu Leu Gln Ser Leu Asn Leu Ala Leu Leu Leu Ala Asp
            350                 355                 360

GTC TGC TTG AAC TTC TTG CCC AGC ATC TAC CTC ATC TTC ATC ATC ATT     1275
Val Cys Leu Asn Phe Leu Pro Ser Ile Tyr Leu Ile Phe Ile Ile Ile
        365                 370                 375

CTG TAC GAA GGG CTC CTG GGT GGG GCC GCT TAC GTG AAT ACC TTC CAC     1323
Leu Tyr Glu Gly Leu Leu Gly Gly Ala Ala Tyr Val Asn Thr Phe His
    380                 385                 390

AAC ATT GCT CTG GAG ACC AGT GAC AAG CAC CGA GAG TTT GCC ATG GAA     1371
Asn Ile Ala Leu Glu Thr Ser Asp Lys His Arg Glu Phe Ala Met Glu
395                 400                 405                 410

GCT GCC TGT ATC TCT GAC ACC TTG GGA ATC TCC CTG TCG GGG GTC CTG     1419
Ala Ala Cys Ile Ser Asp Thr Leu Gly Ile Ser Leu Ser Gly Val Leu
                415                 420                 425

GCC CTG CCT CTG CAT GAC TTC CTC TGT CAC CTC CC TTGACAGGAG           1464
Ala Leu Pro Leu His Asp Phe Leu Cys His Leu
            430                 435
```

TTGCTCGACA CACACTGATC TGCAGGCACA TGAGCAGATC ACACATCTTC GAGCTCTGCC 1524

ACAGCCTTTC CCTGCCCCAC TGCAGCAAGG AGCCCCTGAT GTTTCCCACT CCTGAGCTGG 1584

CCTCAGAGTT TTCTCCTACC CTCTGCCCTT CTAATAAATG CTTATTTTAA CAGTTAAAAA 1644

AAAAAAAAAA AAAA 1658

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Gly Ser Ser Ala Gly Ser Trp Arg Arg Leu Glu Asp Ser Glu Arg
1 5 10 15

Glu Glu Thr Asp Ser Glu Pro Gln Ala Pro Arg Leu Asp Ser Arg Ser
20 25 30

Val Leu Trp Lys Asn Ala Val Gly Phe Trp Ile Leu Gly Leu Cys Asn
35 40 45

Asn Phe Ser Tyr Val Val Met Leu Ser Ala Ala His Asp Ile Leu Lys
50 55 60

Gln Glu Gln Ala Ser Gly Asn Gln Ser His Val Glu Pro Gly Arg Thr
65 70 75 80

Pro Thr Pro His Asn Ser Ser Ser Arg Phe Asp Cys Asn Ser Ile Ser
85 90 95

Thr Ala Ala Val Leu Leu Ala Asp Ile Leu Pro Thr Leu Val Ile Lys
100 105 110

Leu Leu Ala Pro Leu Gly Leu His Leu Leu Pro Tyr Ser Pro Arg Val
115 120 125

Leu Val Ser Gly Val Cys Ser Ala Gly Ser Phe Val Leu Val Ala Phe
130 135 140

Ser Gln Ser Val Gly Leu Ser Leu Cys Gly Val Val Leu Ala Ser Ile
145 150 155 160

Ser Ser Gly Leu Gly Glu Val Thr Phe Leu Ser Leu Thr Ala Phe Tyr
165 170 175

Pro Ser Ala Val Ile Ser Trp Trp Ser Ser Gly Thr Gly Gly Ala Gly
180 185 190

Leu Leu Gly Ser Leu Ser Tyr Leu Gly Leu Thr Gln Ala Gly Leu Ser
195 200 205

```
Pro Gln His Thr Leu Leu Ser Met Leu Gly Ile Pro Val Leu Leu Leu
    210                 215                 220

Ala Ser Tyr Phe Leu Leu Leu Thr Ser Pro Glu Pro Trp Asp Pro Gly
225                 230                 235                 240

Gly Glu Asn Glu Ala Glu Thr Ala Ala Arg Gln Pro Leu Ile Gly Thr
                245                 250                 255

Glu Thr Pro Glu Ser Lys Pro Gly Ala Ser Trp Asp Leu Ser Leu Gln
            260                 265                 270

Glu Arg Trp Thr Val Phe Lys Gly Leu Leu Trp Tyr Ile Ile Pro Leu
        275                 280                 285

Val Leu Val Tyr Phe Ala Glu Tyr Phe Ile Asn Gln Gly Leu Phe Glu
    290                 295                 300

Leu Leu Phe Phe Arg Asn Thr Ser Leu Ser His Ala His Glu Tyr Arg
305                 310                 315                 320

Trp Tyr Gln Met Leu Tyr Gln Ala Gly Val Phe Ala Ser Arg Ser Ser
                325                 330                 335

Leu Gln Cys Cys Arg Ile Arg Phe Thr Trp Val Leu Ala Leu Leu Gln
            340                 345                 350

Ser Leu Asn Leu Ala Leu Leu Leu Ala Asp Val Cys Leu Asn Phe Leu
        355                 360                 365

Pro Ser Ile Tyr Leu Ile Phe Ile Ile Ile Leu Tyr Glu Gly Leu Leu
    370                 375                 380

Gly Gly Ala Ala Tyr Val Asn Thr Phe His Asn Ile Ala Leu Glu Thr
385                 390                 395                 400

Ser Asp Lys His Arg Glu Phe Ala Met Glu Ala Ala Cys Ile Ser Asp
                405                 410                 415

Thr Leu Gly Ile Ser Leu Ser Gly Val Leu Ala Leu Pro Leu His Asp
            420                 425                 430

Phe Leu Cys His Leu
        435
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGGGGAGGAC AAGCACTG 18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CATTCTGTCA CCCTTAGAAG CC 22

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGACTTGAAG GACGGAGTCT 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGAGCCTCTA TGAGCTGATA CTG 23

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTCGTCCTGG TTGCCTTT 18

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCTGATGAGA TGCTAGCGAA 20

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AGACTCCGTC CTTTCAAGTC C 21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTACACATTC GAGGCCAACC T 21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AAAGGTACAG GCCTCAGGGT 20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AGCTCTCATT CCCCTCAGGT 20

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ACCTGAGGGA ATGAGAGCT 19

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TGGGTTCAGC TCCTTTGC 18

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATTGAAGGGC ATAGGTAAGA 20

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACTTTACCCC ACCTTGTCCC 20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TCAAGTGAAG GCAGAGCTGG 20

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AGTCCCAGCT GGGTAGTGAA 20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCTGTGTTTG TAGCAGGCCT 20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AAGGTCGGTC TCTACTCTCA GC 22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TGGTCAGGAG CTGAGAAAGG 20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GAATCCCTTT CCTCTGGGAG 20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGAGCCTCTA TGAGCTGATA CTG 23

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGAACATTCA GGAGGACCTA GG 22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TGTCCCATGG TCAGCCTAG 19

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TTCTCTCCTT GGACCCCTCT 20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GCAGTGAGCT ACCCATCTTT 20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AGGAAAAGGC CAAACCCAG 19

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AATCCAGTGG CATGGAAGTT G 21

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CTACGACCAA GGGAACAAT 19

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CTACGACCAA GGGAACAAT 19

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TCGGGAAAGG TGGACAGT 18

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGTATTGCTG AGCGTGACTC 20

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AGGTGAAACG GATGCGAC 18

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TTTGAACTCC TCTTTTTCTG G 21

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

ACACTTTCCA CTGATAGTGG GA 22

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

TCCTAAAACC AGGGACCCCT 20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TTCAGTCCCA GACATCCCTG 20

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

AGGGATGTCT GGGACTGAAG 20

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GGCATGATGC CAGGAAGA 18

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

AGGAAGGAGG CTGGAGGATA 20

What is claimed is:

1. A substantially pure preparation of a Batten disease polypeptide, said polypeptide having more than 85% homology with an amino acid sequence of SEQ ID NO:2.

2. A substantially pure nucleic acid which encodes a Batten disease polypeptide, said polypeptide having more than 85% homology with an amino acid sequence of SEQ ID NO:2.

3. A probe or primer which comprises a substantially purified oligonucleotide which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of SEQ ID NO:1 or SEQ ID NO:18.

4. The probe or primer of claim 3, wherein said probe or primer is about 10 to 100 nucleotides in length.

5. The probe or primer of claim 3, wherein said probe or primer overlaps the 1.02 Kb deletion of the Batten disease gene.

6. The probe or primer of claim 3, wherein said probe or primer is located inside the 1.02 Kb deletion of the Batten disease gene.

7. The probe or primer of claim 3, wherein said probe or primer is located outside the 1.02 Kb deletion of the Batten disease gene.

8. A method of evaluating whether a mammal is at risk for Batten disease, comprising detecting in a tissue of said mammal the presence or absence of a mutation of a Batten disease gene.

9. The method of claim 4, wherein said detection comprises:

(i) providing a primer which spans the lesion;

(ii) amplifying a nucleic acid of said tissue with said lesion spanning primer; and

(iii) detecting the presence or absence of said lesion.

10. The method of claim 9, wherein said primer overlaps the 1.02 Kb deletion of the Batten disease gene.

11. The method of claim 9, wherein said method further comprises amplifying said nucleic acid with a primer located inside the 1.02 Kb deletion of the Batten disease gene.

12. The method of claim 9, wherein said further comprises amplifying said nucleic acid with a primer located outside the 1.02 Kb deletion of the Batten disease gene.

13. The method of claim 9, wherein said lesion is a deletion in said Batten disease gene.

14. The method of claim 13, wherein said deletion is the1.02 Kb deletion.

15. The method of claim 9, wherein said lesion is selected from the group consisting of a 1 bp deletion, a 2 bp insertion, a nonsense mutation, a missense mutation and a splice site mutation.

16. The method of claim 9, wherein said lesion is selected from those in Table 3.

17. The method of claim 8, wherein said detection comprises sequencing said mutation and comparing a sequence to a wild-type sequence.

18. A method of determining if a subject mammal is at risk for a Batten disease or misexpression of a Batten disease gene, said method comprising detecting in a tissue of said subject misexpression of a Batten disease polypeptide or Batten disease polypeptide RNA.

19. A method of evaluating a compound for the ability to interact with a Batten disease polypeptide, said method comprising contacting said compound with said Batten disease polypeptide and evaluating ability of said compound to interact with said Batten disease polypeptide.

20. A method for evaluating an effect of a treatment used to treat a disorder related to the Batten disease gene, said method comprising administering said treatment to a test cell or an organism and evaluating the effect of said treatment on a parameter related to an aspect of Batten disease.

21. A method of treating a mammal at risk for Batten disease, said method comprising administering to said mammal a therapeutically effective amount of a nucleic acid encoding a Batten disease polypeptide.

22. A method of treating a mammal at risk for Batten disease, said method comprising administering to said mammal a therapeutically effective amount of a Batten disease polypeptide.

23. A transgenic mammal having a Batten disease transgene.

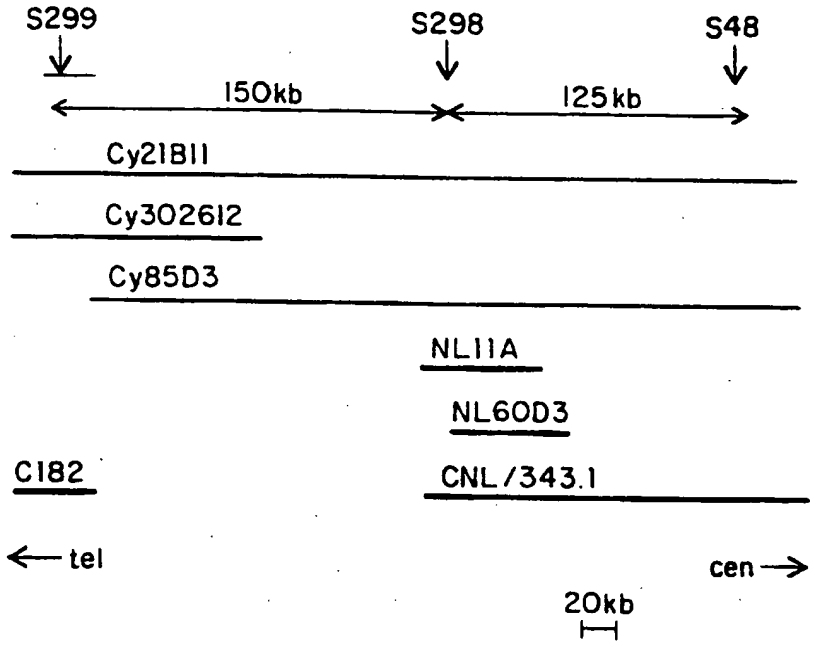
S299
S298
S48
150kb
125kb
Cy21B11
Cy302612
Cy85D3
NL11A
NL60D3
C182
CNL/343.1
tel
cen
20kb

FIG. 1

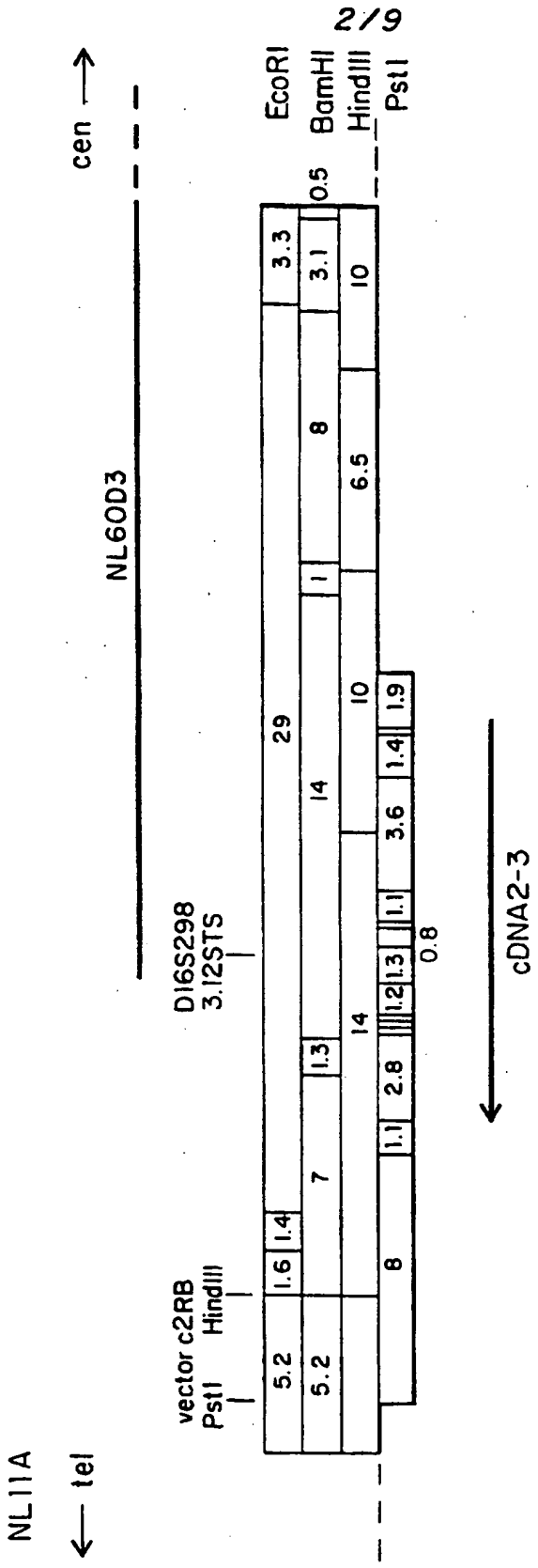
cen
NL60D3
EcoRI
BamHI
HindIII
PstI
0.5
3.3
3.1
10
8
6.5
1
29
14
10
1.9
1.4
3.6
1.1
1.3
1.2
0.8
14
1.3
7
2.8
1.1
8
1.4
1.6
5.2
5.2
D16S298
3.12STS
vector c2RB
PstI HindIII
cDNA2-3
NL11A
tel

FIG. 2

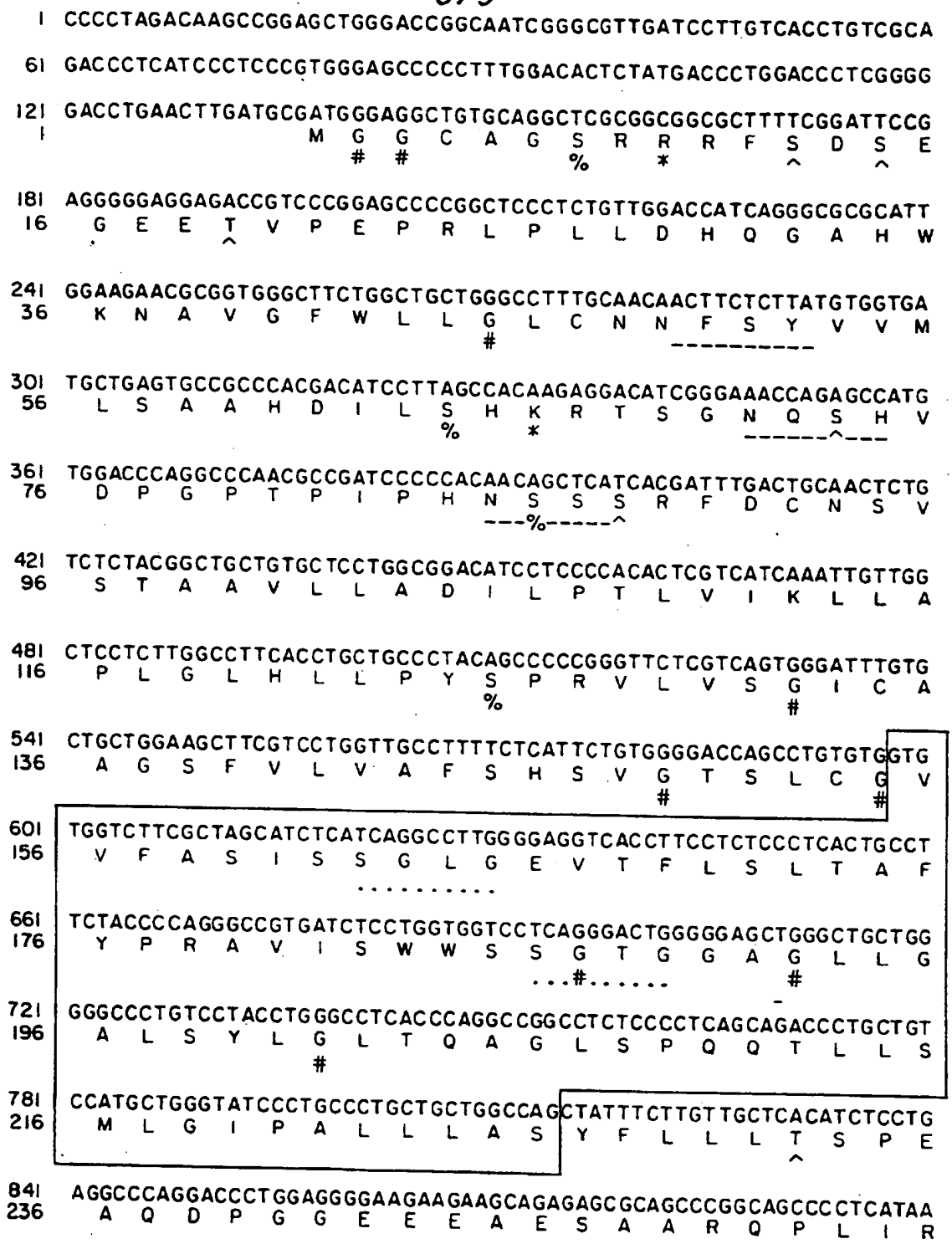
1 CCCCTAGACAAGCCGGAGCTGGGACCGGCAATCGGGCGTTGATCCTTGTCACCTGTCGCA
61 GACCCTCATCCCTCCCGTGGGAGCCCCCTTTGGACACTCTATGACCCTGGACCCTCGGGG
121 GACCTGAACTTGATGCGATGGGAGGCTGTGCAGGCTCGCGGCGGCGCTTTTCGGATTCCG
1 M G G C A G S R R R F S D S E
181 AGGGGGAGGAGACCGTCCCGGAGCCCCGGCTCCCTCTGTTGGACCATCAGGGCGCGCATT
16 G E E T V P E P R L P L L D H Q G A H W
241 GGAAGAACGCGGTGGGCTTCTGGCTGCTGGGCCTTTGCAACAACTTCTCTTATGTGGTGA
36 K N A V G F W L L G L C N N F S Y V V M
301 TGCTGAGTGCCGCCCACGACATCCTTAGCCACAAGAGGACATCGGGAAACCAGAGCCATG
56 L S A A H D I L S H K R T S G N Q S H V
361 TGGACCCAGGCCCAACGCCGATCCCCCACAACAGCTCATCACGATTTGACTGCAACTCTG
76 D P G P T P I P H N S S S R F D C N S V
421 TCTCTACGGCTGCTGTGCTCCTGGCGGACATCCTCCCCACACTCGTCATCAAATTGTTGG
96 S T A A V L L A D I L P T L V I K L L A
481 CTCCTCTTGGCCTTCACCTGCTGCCCTACAGCCCCCGGGTTCTCGTCAGTGGGATTTGTG
116 P L G L H L L P Y S P R V L V S G I C A
541 CTGCTGGAAGCTTCGTCCTGGTTGCCTTTTCTCATTCTGTGGGGACCAGCCTGTGTGGTG
136 A G S F V L V A F S H S V G T S L C G V
601 TGGTCTTCGCTAGCATCTCATCAGGCCTTGGGGAGGTCACCTTCCTCTCCCTCACTGCCT
156 V F A S I S S G L G E V T F L S L T A F
661 TCTACCCCAGGGCCGTGATCTCCTGGTGGTCCTCAGGGACTGGGGGAGCTGGGCTGCTGG
176 Y P R A V I S W W S S G T G G A G L L G
721 GGGCCCTGTCCTACCTGGGCCTCACCCAGGCCGGCCTCTCCCCTCAGCAGACCCTGCTGT
196 A L S Y L G L T Q A G L S P Q Q T L L S
781 CCATGCTGGGTATCCCTGCCCTGCTGCTGGCCAGCTATTTCTTGTTGCTCACATCTCCTG
216 M L G I P A L L L A S Y F L L L T S P E
841 AGGCCCAGGACCCTGGAGGGGAAGAAGAAGCAGAGAGCGCAGCCCGGCAGCCCCTCATAA
236 A Q D P G G E E E A E S A A R Q P L I R

FIG. 3

```
901 GAACCGAGGCCCCGGAGTCGAAGCCAGGCTCCAGCTCCAGCCTCTCCCTTCGGGAAAGGT
256   T  E  A  P  E  S  K  P  G  S  S  S  S  L  S  L  R  E  R  W
                              #                  %^

961 GGACAGTATTCAAGGGTCTGCTGTGGTACATTGTTCCCTTGGTCGTAGTTTACTTTGCCG
276   T  V  F  K  G  L  L  W  Y  I  V  P  L  V  V  V  Y  F  A  E

1021 AGTATTTCATTAACCAGGGACTTTTTGAACTCCTCTTTTTCTGGAACACTTCCCTGAGTC
296   Y  F  I  N  Q  G  L  F  E  L  L  F  F  W  N  T  S  L  S  H
                                                 ----------

1081 ACGCTCAGCAATACCGCTGGTACCAGATGCTGTACCAGGCTGGCGTCTTTGCCTCCCGCT
316   A  Q  Q  Y  R  W  Y  Q  M  L  Y  Q  A  G  V  F  A  S  R  S
                                          #

1141 CTTCTCTCCGCTGCTGTCGCATCCGTTTCACCTGGGCCCTGGCCCTGCTGCAGTGCCTCA
336   S  L  R  C  C  R  I  R  F  T  W  A  L  A  L  L  Q  C  L  N
      %

1201 ACCTGGTGTTCCTGCTGGCAGACGTGTGGTTCGGCTTTCTGCCAAGCATCTACCTCGTCT
356   L  V  F  L  L  A  D  V  W  F  G  F  L  P  S  I  Y  L  V  F

1261 TCCTGATCATTCTGTATGAGGGGCTCCTGGGAGGCGCAGCCTACGTGAACACCTTCCACA
376   L  I  I  L  Y  E  G  L  L  G  G  A  A  Y  V  N  T  F  H  N
                        #

1321 ACATCGCCCTGGAGACCAGTGATGAGCACCGGGAGTTTGCAATGGCGGCCACCTGCATCT
396   I  A  L  E  T  S  D  E  H  R  E  F  A  M  A  A  T  C  I  S
                  ^

1381 CTGACACACTGGGGATCTCCCTGTCGGGGCTCCTGGCTTTGCCTCTGCATGACTTCCTCT
416   D  T  L  G  I  S  L  S  G  L  L  A  L  P  L  H  D  F  L  C
               #

1441 GCCAGCTCTCCTGATACTCGGGATCCTCAGGACGCAGGTCACATTCACCTGTGGGCAGAG
436   Q  L  S

1501 GGACAGGTCAGACACCCAGGCCCACCCCAGAGACCCTCCATGAACTGTGCTCCCAGCCTT

1561 CCCGGCAGGTCTGGGAGTAGGGAAGGGCTGAAGCCTTGTTTCCTTGCAGGGGGGCCAGCC

1621 ATTGTCTCCCACTTGGGGAGTTTCTTCCTGGCATCATGCCTTCTGAATAAATGCCGATTT
                                                    $$$$$$

1681 TGTCCATGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 3
(CONT.)

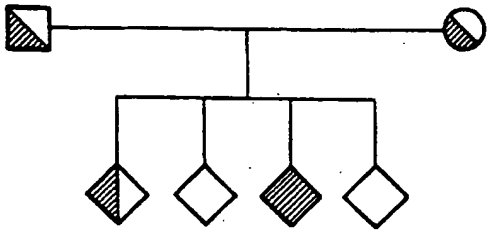

FIG. 4

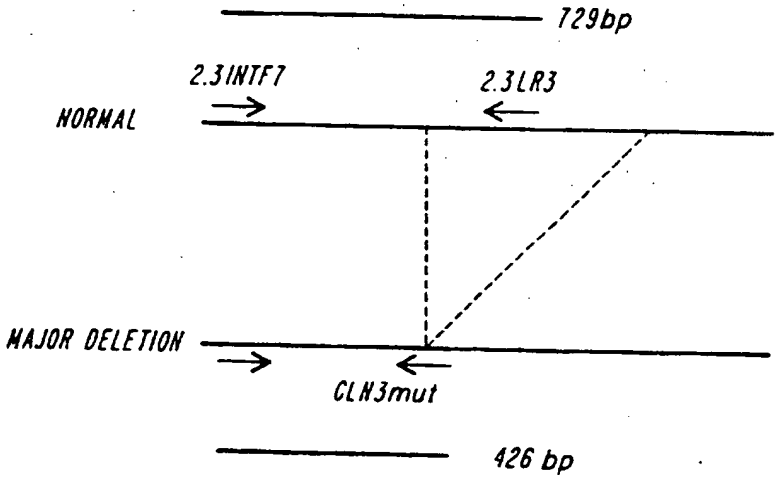

FIG. 7

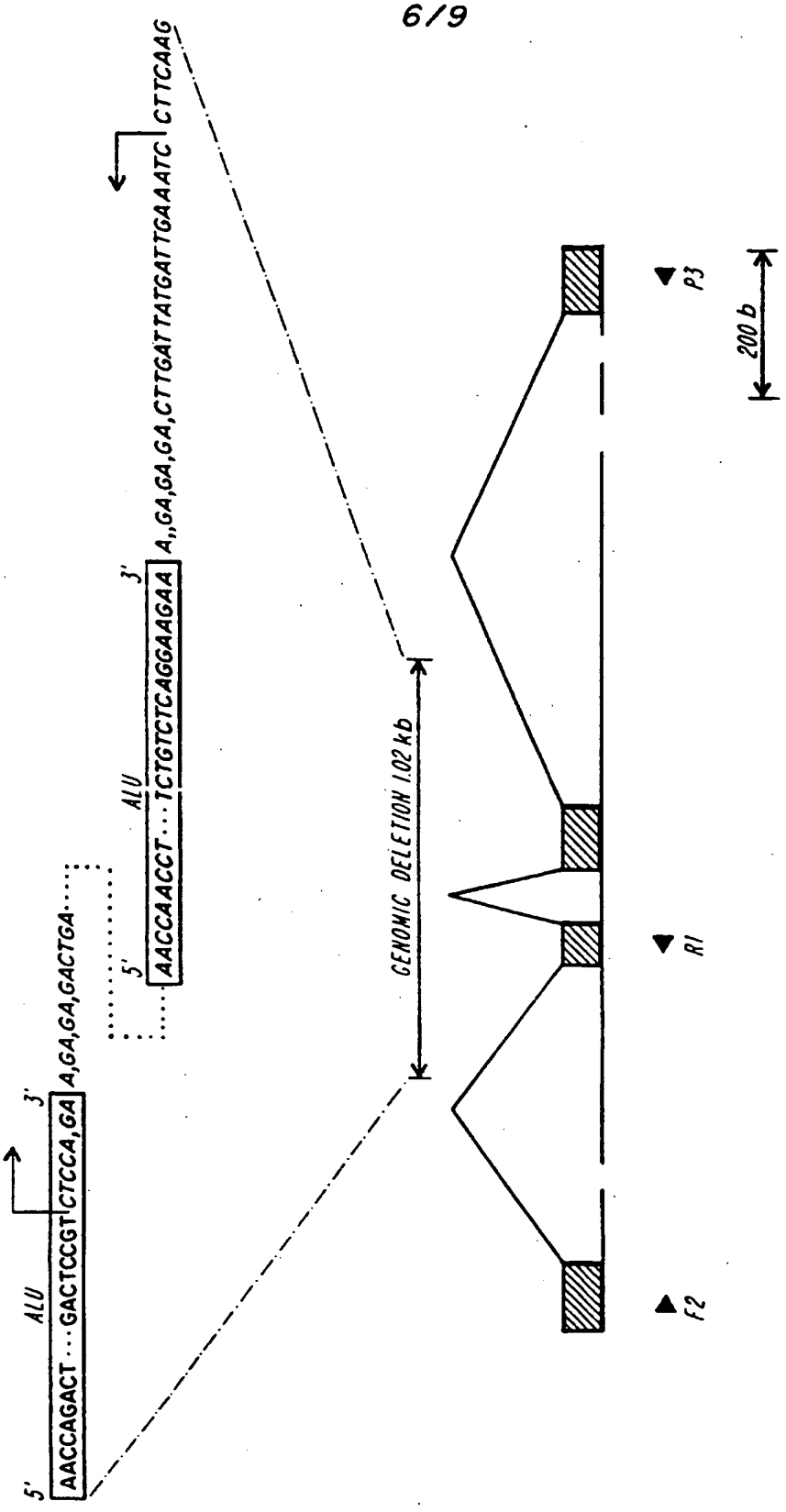
5'
AACCAGACT···GACTCCGT CTCCA,GA
3'
ALU
A,GA,GA,GACTGA
5'
AACCAACCT···TCTGTCTCAGGAAGAA
3'
ALU
A,,GA,GA,GA,CTTGATTATGATTGAAATC CTTCAAG
GENOMIC DELETION 1.02 kb
P3
200 b
R1
F2

FIG. 5

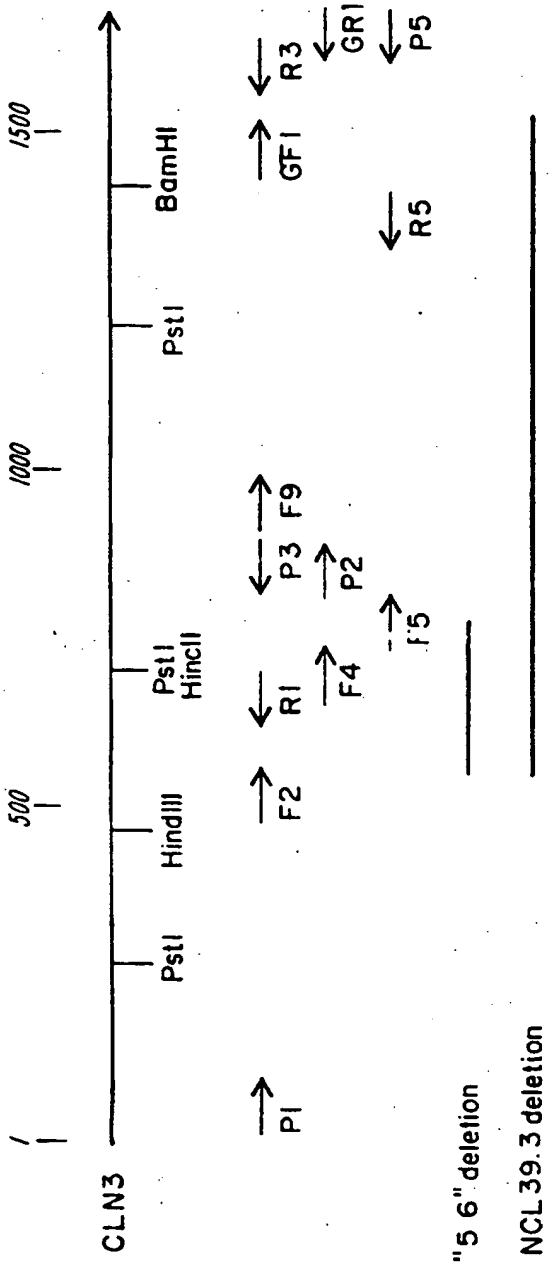
1
500
1000
1500
CLN3
PstI
HindIII
PstI
HincII
PstI
BamHI
P1
F2
R1
F4
P3
P2
F5
F9
GF1
R3
GR1
P5
R5
"5 6" deletion
NCL 39.3 deletion

FIG. 6

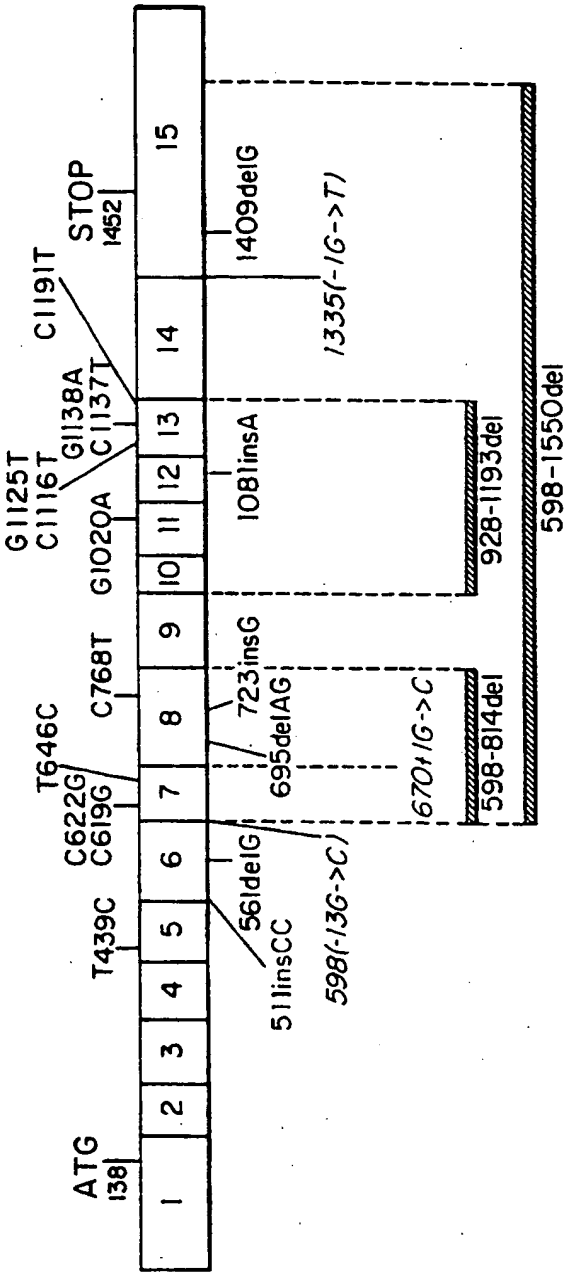
ATG
138
STOP
1452
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
T439C
C622G
C619G
T646C
C768T
G1020A
G1125T
C1116T
G1138A
C1137T
C1191T
511insCC
561delG
598(-13G->C)
695delAG
723insG
1081insA
1335(-1G->T)
1409delG
598-814del
928-1193del
598-1550del
670(1G->C)

FIG. 8

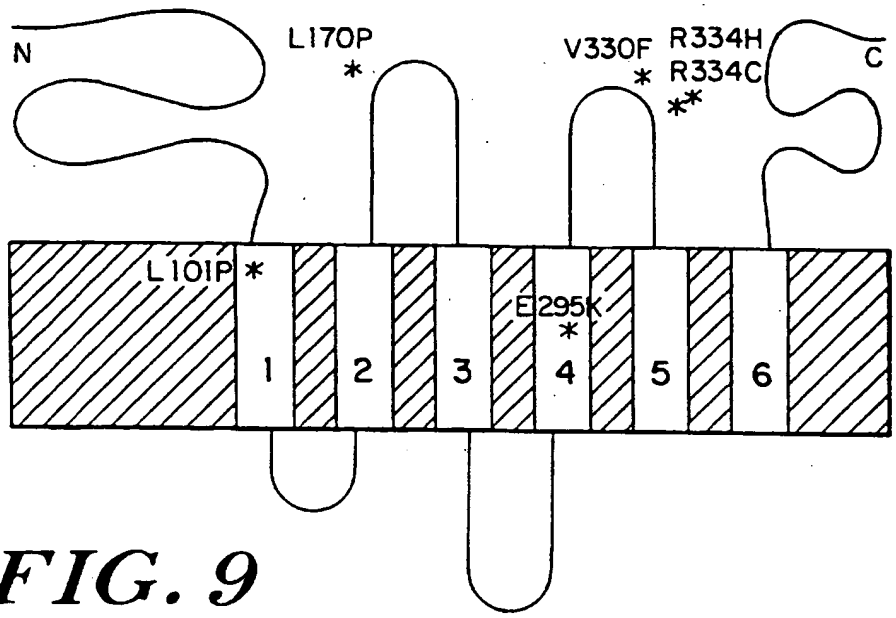
N
L170P
V330F
R334H
R334C
C
L101P
E295K
1
2
3
4
5
6

FIG. 9

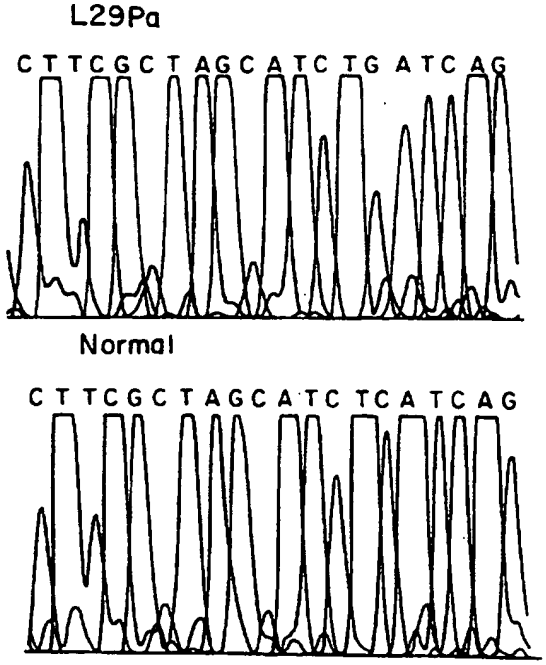
L29Pa
CTTCGCTAGCATCTGATCAG
Normal
CTTCGCTAGCATCTCATCAG

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/13896

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :536/23.1, 24.3; 530/350; 435/6; 514/44; 800/2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.3; 530/350; 435/6; 514/44; 800/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, EMBASE, BIOSIS, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P ------ Y, P	LERNER et al. Isolation of a Novel Gene Underlying Batten Disease, CLN3. Cell. 22 September 1995, Vol. 82, pages 949-957, see entire document.	1-7 ------ 8-23
Y	LERNER et al. Isolation of Genes From the Batten Candidate Region Using Exon Amplification. American Journal of Medical Genetics. 1995, Vol. 57, pages 320-323, see entire document.	1-23
Y	JARVELA et al. Physical Map of the Region Containing the Gene for Batten Disease (CLN3). American Journal of Medical Genetics. 1995, Vol. 57, pages 316-319, see entire document.	1-23

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date		
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 14 NOVEMBER 1996	Date of mailing of the international search report 03 DEC 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer D. CURTIS HOGUE, JR. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/13896

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MITCHISON et al. Refined Localization of the Batten Disease Gene (CLN3) by Haplotype and Linkage Disequilibrium Mapping to D16S288-D16S383 and Exclusion From This Region of a Variant Form of Batten Disease With Granular Osmiophilic Deposits. American Journal of Medical Genetics. 1995, Vol. 57, pages 312-315, see entire document.	1-23
Y	MITCHISON et al. Genetic Mapping of the Batten Disease Locus (CLN3) to the Interval D16S288-D16S383 by Analysis of Haplotypes and Allelic Association. Genomics. 1994, Vol. 22, pages 465-468, see entire document.	1-23
Y	LERNER et al. Linkage Disequilibrium between the Juvenile Neuronal Ceroid Lipofuscinosis Gene and Marker Loci on Chromosome 16p12.1. American Journal of Human Genetics. 1994, Vol. 54, pages 88-94, see entire document.	1-23
Y	LERNER et al. Isolation of candidate genes from the Batten Disease (JNCL) region using exon amplification. American Journal of Human Genetics. 18-22 October 1994, Vol. 55, No. 3, page A263, abstract no. 1541, see entire abstract.	1-23

INTERNATIONAL SEARCH REPORT

Inter...ational application No.
PCT/US96/13896

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12N 15/11; C07K 14/00; C07H 21/04; C12Q 1/68; A61K 48/00; A01K 67/00